118

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 January 2001 (25.01.2001)

PCT

(10) International Publication Number
WO 01/05970 A2

(51) International Patent Classification[7]: C12N 15/12, C07K 14/47, G01N 33/53, C12Q 1/68, A61K 38/17, C07K 16/18, A01K 67/027

(21) International Application Number: PCT/US00/19698

(22) International Filing Date: 19 July 2000 (19.07.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/144,595	19 July 1999 (19.07.1999)	US
60/150,460	23 August 1999 (23.08.1999)	US
60/159,849	15 October 1999 (15.10.1999)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/144,595 (CIP)
Filed on	19 July 1999 (19.07.1999)
US	60/150,460 (CIP)
Filed on	23 August 1999 (23.08.1999)
US	60/159,849 (CIP)
Filed on	15 October 1999 (15.10.1999)

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **TANG, Y., Tom** [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **BANDMAN, Olga** [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). **HILLMAN, Jennifer, L.** [US/US]; 230 Monroe Drive #12, Montain View, CA 94040 (US). **LAL, Preeti** [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). **AU-YOUNG, Janice** [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). **REDDY, Roopa** [IN/US]; 1233 W. McKinley Avenue, #3, Sunnyvale, CA 94086 (US). **YANG, Junming** [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **LU, Dyung, Aina, M.** [US/US]; 55 Park Belmont Place, San Jose, CA 95136 (US). **AZIMZAI, Yalda** [US/US]; 2045 Rock Springs Drive, Hayward, CA 94545 (US). **PATTERSON, Chandra** [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US).

(74) Agents: **HAMLET-COX, Diana et al.**; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/05970 A2

(54) Title: GTP-BINDING ASSOCIATED PROTEINS

(57) Abstract: The invention provides human GTP-binding associated proteins (GBAP) and polynucleotides which identify and encode GBAP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of GBAP.

GTP-BINDING ASSOCIATED PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of GTP-binding associated proteins and to the use of these sequences in the diagnosis, treatment, and prevention of immune system, reproductive, nervous system, and cell signaling disorders, and cell proliferative disorders including cancer.

BACKGROUND OF THE INVENTION

Guanine nucleotide binding proteins (GTP-binding proteins) are present in all eukaryotic cells and function in processes including metabolism, cellular growth, differentiation, signal transduction, cytoskeletal organization, and intracellular vesicle transport and secretion. In higher organisms they are involved in signaling that regulates such processes as the immune response (Aussel, C. et al. (1988) J. Immunol. 140:215-220), apoptosis, differentiation, and cell proliferation including oncogenesis (Dhanasekaran, N. et al. (1998) Oncogene 17:1383-1394).

The superfamily of GTP-binding proteins can be subdivided into groups such as translational factors, heterotrimeric GTP-binding proteins involved in transmembrane signaling processes (also called G-proteins), proto-oncogene Ras proteins, other low molecular weight GTP-binding proteins including the products of rab, rap, rho, rac, smg21, smg25, YPT, SEC4, and ARF genes, and tubulins (Kaziro, Y. et al. (1991) Annu. Rev. Biochem. 60:349-400).

GTP-binding proteins are involved in protein biosynthesis and include initiation factor 2 (IF-2), elongation factor 2 (EF-Tu), and elongation factor G (EF-G), observed in prokaryotes; and initiation factor 2 (eIF-2), elongation factor Iα (EF-Iα), elongation factor 2 (EF-2), and release factor 3 (eRF3) observed in eukaryotes (Kaziro, supra). IF-2 promotes the GTP-dependent binding of the tRNA to the small subunit of the ribosome, the step that initiates protein translation. Elongation factors promote the binding of tRNA and GTP and the displacement of GDP after hydrolysis as protein biosynthesis proceeds. eRF3 participates in the recognition of stop codons and the release of nascent proteins from ribosomes.

Heterotrimeric GTP-binding proteins are composed of 3 subunits (α, β and γ) which, in the resting state, associate as a trimer at the inner face of the plasma membrane. Heterotrimeric G-proteins may be classified based on the sequence similarity of α subunits into the Gs, Gi, Gq and G12 subgroups. In the resting state, the α subunit binds guanosine diphosphate (GDP), and stimulation of the G-protein by an activated receptor leads to exchange of GDP for guanosine triphosphate (GTP). This exchange results in the separation of the α from the β and γ subunits, which remain tightly

associated as a dimer. Both the α subunit and β-γ dimer are then able to interact with effectors, either individually or in a cooperative manner. The intrinsic GTPase activity of the α subunit hydrolyzes the bound GTP to GDP. This returns the α subunit to its inactive conformation and allows it to reassociate with the β-γ complex, thus restoring the system to its resting state (Kaziro, supra). Some α subunits show tissue-specific expression indicating a specialized signaling role (Dhanasekaran, supra).

The α-s class of G-protein subunits is sensitive to ADP-ribosylation by pertussis toxin which uncouples the receptor:G-protein interaction. This uncoupling blocks signal transduction to receptors that decrease cAMP levels. cAMP levels regulate ion channels and activate phospholipases. The inhibitory α-I class is also susceptible to modification by pertussis toxin, which prevents α-I from lowering cAMP levels. Two novel classes of α subunits refractory to pertussis toxin modification are α-q, which activates phospholipase C, and α-12, which has sequence homology with the Drosophila gene concertina and may contribute to the regulation of embryonic development (Simon, M.I. (1991) Science 252:802-808).

The mammalian G-protein β and γ subunits, each about 340 amino acids long, share more than 80% homology. The β subunit (also called β-transducin) contains seven repeating units, each about 43 amino acids long. This WD-repeat, or G-beta repeat motif, is found in a variety of proteins with regulatory function such as Sec13, a yeast WD repeat protein involved in vesicular traffic; coronin-2, a mammalian WD repeat protein involved in regulation of the actin cytoskeleton; and Bop1, a mammalian WD repeat protein involved in growth suppression (Garcia-Higuera, I. et al. (1998) J. Biol. Chem. 273:9041-9049; Okumura, M. et al. (1998) DNA Cell Biol. 17:779-787; Pestov, D.G. et al. (1998) Oncogene 17:3187-3197). The activity of the β and γ subunits may be regulated by other proteins such as calmodulin, phosducin, or the neural protein GAP 43 (Clapham, D.E. and E.J. Neer (1993) Nature 365:403-406). The β subunit sequences are highly conserved among species, suggesting that they perform a fundamentally important role in the organization and function of G-protein linked systems (Van der Voorn, L. and H.L. Ploegh (1992) FEBS Lett. 307:131-134).

Mutations and variant expression of β-transducin proteins are linked with various disorders. Mutations in LIS1, a subunit of the human platelet activating factor acetylhydrolase, cause Miller-Dieker lissencephaly. RACK1 binds activated protein kinase C, and RbAp48 binds retinoblastoma protein. CstF is required for polyadenylation of mammalian pre-mRNA in vitro and associates with subunits of cleavage-stimulating factor. Defects in the regulation of β-catenin contribute to the neoplastic transformation of human cells. The WD40 repeats of the human F-box protein βTrCP mediate binding to β-catenin, thus regulating the targeted degradation of β-catenin by ubiquitin ligase (Neer, E.J. et al. (1994) Nature 371:297-300; Hart, M. et al. (1999) Curr. Biol. 9:207-210).

The γ subunit sequences are more variable than those of the β subunits. They are often post-translationally modified by isoprenylation and carboxyl-methylation of a cysteine residue four amino

acids from the C-terminus. These modifications appear to be necessary for the interaction of the β-γ dimer with the membrane and with other GTP-binding proteins. The β-γ dimer has been shown to modulate the activity of adenylyl cyclase isoforms, phospholipase C, and some ion channels. It is involved in receptor phosphorylation via specific kinases and has been implicated in the p21ras-dependent activation of the MAP kinase cascade and the recognition of specific receptors by GTP-binding proteins (Clapham and Neer, supra).

G-proteins interact with a variety of effectors including adenylyl cyclase (Clapham and Neer, supra). The signaling pathway mediated by cAMP is mitogenic in hormone-dependent endocrine tissues such as adrenal cortex, thyroid, ovary, pituitary, and testes. Cancers in these tissues have been related to a mutationally activated form of a $G\alpha_s$ known as the gsp (Gs protein) oncogene (Dhanasekaran, supra). Another effector is phosducin, a retinal phosphoprotein, which forms a specific complex with retinal G-protein β and γ subunits and modulates the ability of the β-γ dimer to interact with retinal α subunits (Clapham and Neer, supra). Additional G-protein effectors include RIN1 (Ras interaction/interference), which acts as an effector of H-Ras and interferes with the Ras signal transduction pathway; Rabin3, which associates with the Ras-like GTPase Rab3A; and Rhotekin, a protein that binds with, and inhibits, Rho GTPase activity (Han, L. and J. Colicelli (1995) Mol. Cell Biol. 15:1318-1323; Brondyk, W.H. et al. (1995) Mol. Cell Biol. 15:1137-1143; and Reid, T. et al. (1996) J. Biol. Chem. 27:13556-13560).

The low molecular weight GTP-binding proteins regulate cell growth, cell cycle control, protein secretion, and intracellular vesicle interaction. These GTP-binding proteins respond to extracellular signals from receptors and activating proteins by transducing mitogenic signals (Tavitian, A. (1995) C. R. Seances Soc. Biol. Fil. 189:7-12). Low molecular weight GTP-binding proteins consist of single polypeptides of 21-30kD which, like the α subunit of heterotrimeric GTP-binding proteins, are able to bind to and hydrolyze GTP, thus cycling from an inactive to an active state. The intrinsic rate of GTP hydrolysis of these GTP-binding proteins is typically very slow, but it can be stimulated by several orders of magnitude by GTPase-activating proteins (GAPs), such as β2-chimaerin (Geyer, M. and Wittinghofer, A. (1997) Curr. Opin. Struct. Biol. 7:786-792; Caloca, M. J. et al. (1997) J. Biol. Chem. 272:26488-26496).

Low molecular weight GTP-binding proteins play critical roles in cellular protein trafficking events, such as the translocation of proteins and soluble complexes from the cytosol to the membrane through an exchange of GDP for GTP (Ktistakis, N.T. (1998) BioEssays 20:495-504). In vesicle transport, the interaction between vesicle- and target- specific identifiers (v-SNAREs and tSNAREs) docks the vesicle to the acceptor membrane. The budding process is regulated by GTPases such as the closely related ADP ribosylation factors (ARFs) and SAR proteins, while GTPases such as Rab allow assembly of SNARE complexes and may play a role in removal of defective complexes (Rothman, J.E.

and F.T. Wieland (1996) Science 272:227-234). The rab proteins control the translocation of vesicles to and from membranes for protein localization, protein processing, and secretion. The rho GTP-binding proteins control signal transduction pathways that link growth factor receptors to actin polymerization which is necessary for normal cellular growth and division. The ran GTP-binding proteins are located in the nucleus of cells and have a key role in nuclear protein import, the control of DNA synthesis, and cell-cycle progression (Hall, A. (1990) Science 249:635-640; Scheffzek, K. et al. (1995) Nature 374:378-381).

The Ras proteins Ras1, Ras2 and $G_s\alpha$ stimulate adenylyl cyclase (Kaziro, supra) which affects a broad array of cellular processes including determination of whether cells continue to grow or become terminally differentiated. Stimulation of cell surface receptors activates Ras which, in turn, activates cytoplasmic kinases. These kinases translocate to the nucleus and activate key transcription factors that control gene expression and protein synthesis (Barbacid, M. (1987) Annu. Rev. Biochem. 56:779-827; Treisman, R. (1994) Curr. Opin. Genet. Dev. 4:96-101). Mutant Ras-family proteins which bind but cannot hydrolyze GTP are permanently activated and are thus rendered oncogenic (Drivas, G.T. et al. (1990) Mol. Cell. Biol. 10:1793-1798).

Ras-like proteins have also been implicated in tumor suppression. For example, NOEY2, a novel gene encoding a Ras-like protein, is expressed in normal ovarian and breast epithelial cells. However, NOEY2 expression is reduced or abrogated in ovarian and breast carcinomas, suggesting a role for the NOEY2 gene product in tumor suppression (Yu, Y. et al. (1999) Proc. Natl. Acad. Sci. USA 96:214-219).

Irregularities in GTP-binding protein signaling cascades may result in abnormal activation of leukocytes and lymphocytes, leading to the tissue damage and destruction seen in many inflammatory and autoimmune diseases such as rheumatoid arthritis, biliary cirrhosis, hemolytic anemia, lupus erythematosus, and thyroiditis. Abnormal cell proliferation, including cyclic AMP-mediated stimulation of brain, thyroid, adrenal, and gonadal tissue proliferation is regulated by G proteins. Mutations in G_α subunits have been found in growth-hormone-secreting pituitary somatotroph tumors, hyperfunctioning thyroid adenomas, and ovarian and adrenal neoplasms (Meij, J.T.A. (1996) Mol. Cell. Biochem. 157:31-38; Aussel, supra).

The discovery of new GTP-binding associated proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of immune system, reproductive, nervous system, and cell signaling disorders, and cell proliferative disorders including cancer.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, GTP-binding associated proteins, referred to

collectively as "GBAP" and individually as "GBAP-1," "GBAP-2," "GBAP-3," "GBAP-4," "GBAP-5," "GBAP-6," "GBAP-7," "GBAP-8," "GBAP-9," "GBAP-10," "GBAP-11," "GBAP-12," "GBAP-13," "GBAP-14," "GBAP-15," "GBAP-16," "GBAP-17," "GBAP-18," "GBAP-19," "GBAP-20," "GBAP-21," "GBAP-22," "GBAP-23," "GBAP-24," "GBAP-25," "GBAP-26," "GBAP-27," "GBAP-28," "GBAP-29," "GBAP-30," "GBAP-31," "GBAP-32," "GBAP-33," "GBAP-34," "GBAP-35," "GBAP-36," "GBAP-37," "GBAP-38," "GBAP-39," "GBAP-40," "GBAP-41," "GBAP-42," "GBAP-43," "GBAP-44," "GBAP-45," "GBAP-46," "GBAP-47," "GBAP-48," "GBAP-49," "GBAP-50," "GBAP-51," "GBAP-52," "GBAP-53," "GBAP-54," "GBAP-55," "GBAP-56," "GBAP-57," "GBAP-58," "GBAP-59," "GBAP-60," "GBAP-61," "GBAP-62," "GBAP-63," "GBAP-64," "GBAP-65," and "GBAP-66." In one aspect, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-66. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-66.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-66. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-66. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:67-132.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-66. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism

comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-66. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-66.

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:67-132, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:67-132, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:67-132, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:67-132, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or

fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:67-132, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:67-132, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, and a pharmaceutically acceptable excipient. In one embodiment, the 1 composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-66. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional GBAP, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-66. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of

functional GBAP, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-66. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional GBAP, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-66. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-66, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-66. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a

change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:67-132, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:67-132, ii) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:67-132, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:67-132, ii) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:67-132, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of the above polynucleotide sequence; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 shows polypeptide and nucleotide sequence identification numbers (SEQ ID NOs), clone identification numbers (clone IDs), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding GBAP.

Table 2 shows features of each polypeptide sequence, including potential motifs, homologous sequences, and methods, algorithms, and searchable databases used for analysis of GBAP.

Table 3 shows selected fragments of each nucleic acid sequence; the tissue-specific expression

patterns of each nucleic acid sequence as determined by northern analysis; diseases, disorders, or conditions associated with these tissues; and the vector into which each cDNA was cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which cDNA clones encoding GBAP were isolated.

Table 5 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"GBAP" refers to the amino acid sequences of substantially purified GBAP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of GBAP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of GBAP either by directly interacting with GBAP or by acting on components of the biological pathway in which GBAP participates.

An "allelic variant" is an alternative form of the gene encoding GBAP. Allelic variants may

result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding GBAP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as GBAP or a polypeptide with at least one functional characteristic of GBAP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding GBAP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding GBAP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent GBAP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of GBAP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of GBAP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of GBAP either by directly interacting with GBAP or by acting on components of the biological pathway in which GBAP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind GBAP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic GBAP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or

amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding GBAP or fragments of GBAP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (PE Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the

side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

A "fragment" is a unique portion of GBAP or the polynucleotide encoding GBAP which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50% of a polypeptide) as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:67-132 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:67-132, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:67-132 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:67-132 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:67-132 and the region of SEQ ID NO:67-132 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-66 is encoded by a fragment of SEQ ID NO:67-132. A fragment of SEQ ID NO:1-66 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-66. For example, a fragment of SEQ ID NO:1-66 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-66. The precise length of a fragment of SEQ ID NO:1-66 and the region of SEQ ID NO:1-66 to which the

fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full-length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full-length" polynucleotide sequence encodes a "full-length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (Apr-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions

for nucleic acid hybridization are well known and can be found in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of GBAP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of GBAP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of GBAP. For example, modulation

may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of GBAP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an GBAP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of GBAP.

"Probe" refers to nucleic acid sequences encoding GBAP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for

example Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al.,1987, Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al., 1990, PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a

recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding GBAP, or fragments thereof, or GBAP itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters,

chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants, and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides

due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human GTP-binding associated proteins (GBAP), the polynucleotides encoding GBAP, and the use of these compositions for the diagnosis, treatment, or prevention of immune system, reproductive, nervous system, and cell signaling disorders, and cell proliferative disorders including cancer.

Table 1 lists the Incyte clones used to assemble full length nucleotide sequences encoding GBAP. Columns 1 and 2 show the sequence identification numbers (SEQ ID NOs) of the polypeptide and nucleotide sequences, respectively. Column 3 shows the clone IDs of the Incyte clones in which nucleic acids encoding each GBAP were identified, and column 4 shows the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones and their corresponding cDNA libraries. Clones for which cDNA libraries are not indicated were derived from pooled cDNA libraries. In some cases, GenBank sequence identifiers are also shown in column 5. The Incyte clones and GenBank cDNA sequences, where indicated, in column 5 were used to assemble the consensus nucleotide sequence of each GBAP and are useful as fragments in hybridization technologies.

The columns of Table 2 show various properties of each of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3 shows potential phosphorylation sites; column 4 shows potential glycosylation sites; column 5 shows the amino acid residues comprising signature sequences and motifs; column 6 shows homologous sequences as identified by BLAST analysis; and column 7 shows analytical methods and in some cases, searchable databases to which the analytical methods were applied. The methods of column 7 were used to characterize each polypeptide through sequence homology and protein motifs.

The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding GBAP. The first column of Table 3 lists the nucleotide SEQ ID NOs. Column 2 lists fragments of the nucleotide sequences of column 1. These fragments are useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:67-132 and to distinguish between SEQ ID NO:67-132 and related polynucleotide sequences. The polypeptides encoded by these fragments are useful, for example, as immunogenic peptides. Column 3 lists tissue categories which express GBAP as a fraction of total tissues expressing GBAP. Column 4 lists diseases, disorders, or conditions associated with those tissues expressing GBAP as a fraction of total tissues expressing GBAP. Column 5 lists the vectors used to subclone each cDNA library. Of particular note is the expression of SEQ ID NO:84 in lung tissues, and the tissue-specific expression of SEQ ID NO:132. Over 90% of tissues expressing SEQ ID NO:132 are derived from the nervous system, particularly the brain.

The columns of Table 4 show descriptions of the tissues used to construct the cDNA libraries from which cDNA clones encoding GBAP were isolated. Column 1 references the nucleotide SEQ ID NOs, column 2 shows the cDNA libraries from which these clones were isolated, and column 3 shows the tissue origins and other descriptive information relevant to the cDNA libraries in column 2.

SEQ ID NO:70 maps to chromosome 7 within the interval from 111.6 to 123.4 centiMorgans. This interval contains a gene that is down regulated in adenoma. SEQ ID NO:74 maps to chromosome 11 within the interval from 104.8 to 123.5 centiMorgans. This interval contains a gene associated with the cerebellar degenerative disorder, ataxia telangiectasia. SEQ ID NO:75 maps to chromosome 17 within the interval from 62.9 to 65.0 centiMorgans. SEQ ID NO:77 maps to chromosome 3 within the interval from 12.9 to 16.5 centiMorgans. SEQ ID NO:80 maps to chromosome 9 within the interval from 42.0 to 57.3 centiMorgans. SEQ ID NO:86 maps to chromosome 1 within the interval from 159.6 to 164.1 centiMorgans. SEQ ID NO:87 maps to chromosome 11 within the interval from 147.2 to 151.6. SEQ ID NO:90 maps to chromosome 1 within the interval from 219.2 to 223.0 centiMorgans. This interval contains a gene encoding a RAB interacting protein. SEQ ID NO:92 and SEQ ID NO:106 both map to chromosome 1 within the interval from 48.8 to 81.6 centiMorgans. This interval also contains genes associated with familial hypercholesterolemia, glucose transport defect, infantile hypophosphatasia, infantile neuronal ceroid lipofuscinosis, Kostmann disease, multiple epiphyseal dysplasia, porphyria cutanea tarda, and T-cell acute lymphocytic leukemia 1. SEQ ID NO:93 maps to chromosome 12 within the interval from 76.5 to 87.6 centiMorgans. This interval also contains genes associated with mucopolysaccharidosis type IIID, pseudovitamin D deficiency rickets, and renal amyloidosis. SEQ ID NO:94 and SEQ ID NO:109 both map to chromosome 1 within the interval from 143.1 to 146.6 centiMorgans, to chromosome 14 within the interval from 46.8 to 50.9 centiMorgans, to chromosome 16 within the interval from 88.1 to 90.2 centiMorgans, and to chromosome 19 within the

interval from 58.7 to 97.5 centiMorgans. The interval on chromosome 14 from 46.8 to 50.9 centiMorgans also contains a gene associated with dopa-responsive dystonia. The interval on chromosome 19 from 58.7 to 97.5 centiMorgans also contains genes associated with colorectal cancer, DNA ligase I deficiency, glutaricaciduria IIB, myotonic dystrophy, renal amyloidosis, T-cell acute lymphoblastic leukemia, and xeroderma pigmentosum D. SEQ ID NO:97 maps to chromosome 2 within the interval from 236.2 to 269.5 centiMorgans. This interval also contains genes associated with Crigler-Najjar syndrome, familial hypercholesterolemia, Oguchi disease, and primary hyperoxaluria. SEQ ID NO:101 maps to chromosome 2 within the interval from 225.6 to 233.1 centiMorgans, to chromosome 6 within the interval from 132.7 to 144.4 centiMorgans, and to chromosome 11 within the interval from 117.9 to 120.8 centiMorgans. The interval on chromosome 2 from 225.6 to 233.1 centiMorgans also contains a gene associated with Waardenburg syndrome 1. The interval on chromosome 6 from 132.7 to 144.4 centiMorgans also contains genes associated with familial disseminated atypical mycobacterial infection and rhizomelic chondrodysplasia punctata. The interval on chromosome 11 from 117.9 to 120.8 centiMorgans also contains a gene associated with acute intermittent porphyria. SEQ ID NO:111 maps to chromosome 19 within the interval from 35.5 to 49.4 centiMorgans, to chromosome 1 within the interval from the p-terminus to 16.4 centiMorgans, and to chromosome 11 within the interval from 147.2 centiMorgans to the q-terminus. SEQ ID NO:112 maps to chromosome 19 within the interval from 41.7 to 49.4 centiMorgans. SEQ ID NO:113 maps to chromosome 9 within the interval from 136.2 to 163.0 centiMorgans. SEQ ID NO:115 maps to chromosome 14 within the interval from 95.5 to 103.7 centiMorgans and to the X chromosome (23) within the interval from the p-terminus to 55.5 centiMorgans. SEQ ID NO:117 maps to chromosome 13 at 46.9 centiMorgans. SEQ ID NO:118 maps to chromosome 1 within the interval from 16.4 to 22.9 centiMorgans. SEQ ID NO:121 maps to chromosome 12 within the interval from 116.6 to 118.9 centiMorgans. SEQ ID NO:128 maps to chromosome 1 within the interval from the p-terminus to 16.4 centiMorgans.

The invention also encompasses GBAP variants. A preferred GBAP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the GBAP amino acid sequence, and which contains at least one functional or structural characteristic of GBAP.

The invention also encompasses polynucleotides which encode GBAP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:67-132, which encodes GBAP. The polynucleotide sequences of SEQ ID NO:67-132, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding GBAP. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding GBAP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:67-132 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:67-132. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of GBAP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding GBAP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring GBAP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode GBAP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring GBAP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding GBAP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding GBAP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode GBAP and GBAP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding GBAP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:67-132 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol.

152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (PE Biosystems, Foster City CA), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (PE Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (PE Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding GBAP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a

GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, PE Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode GBAP may be cloned in recombinant DNA molecules that direct expression of GBAP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express GBAP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter GBAP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of GBAP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then

subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding GBAP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, GBAP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (PE Biosystems). Additionally, the amino acid sequence of GBAP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active GBAP, the nucleotide sequences encoding GBAP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding GBAP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding GBAP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding GBAP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous

translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding GBAP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding GBAP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al., (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding GBAP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding GBAP can be achieved using a

multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding GBAP into the vector's multiple cloning site disrupts the *lac*Z gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of GBAP are needed, e.g. for the production of antibodies, vectors which direct high level expression of GBAP may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of GBAP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, supra; and Scorer, supra.)

Plant systems may also be used for expression of GBAP. Transcription of sequences encoding GBAP may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, supra; Broglie, supra; and Winter, supra.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding GBAP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses GBAP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of GBAP in cell lines is preferred. For example, sequences encoding GBAP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*$^-$ and *apr*$^-$ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding GBAP is inserted within a marker gene sequence, transformed cells containing sequences encoding GBAP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding GBAP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding GBAP and that express GBAP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based

technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of GBAP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on GBAP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding GBAP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding GBAP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding GBAP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode GBAP may be designed to contain signal sequences which direct secretion of GBAP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture

Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding GBAP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric GBAP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of GBAP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the GBAP encoding sequence and the heterologous protein sequence, so that GBAP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled GBAP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

GBAP of the present invention or fragments thereof may be used to screen for compounds that specifically bind to GBAP. At least one and up to a plurality of test compounds may be screened for specific binding to GBAP. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of GBAP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which GBAP binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express GBAP, either as a secreted

protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing GBAP or cell membrane fractions which contain GBAP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either GBAP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with GBAP, either in solution or affixed to a solid support, and detecting the binding of GBAP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

GBAP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of GBAP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for GBAP activity, wherein GBAP is combined with at least one test compound, and the activity of GBAP in the presence of a test compound is compared with the activity of GBAP in the absence of the test compound. A change in the activity of GBAP in the presence of the test compound is indicative of a compound that modulates the activity of GBAP. Alternatively, a test compound is combined with an in vitro or cell-free system comprising GBAP under conditions suitable for GBAP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of GBAP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding GBAP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred

to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding GBAP may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding GBAP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding GBAP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress GBAP, e.g., by secreting GBAP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of GBAP and GTP-binding associated proteins. In addition, the expression of GBAP is closely associated with reproductive tissues, inflammation and the immune response, trauma, cell proliferation, and cancer. Therefore, GBAP appears to play a role in immune system, reproductive, nervous system, and cell signaling disorders, and cell proliferative disorders including cancer. In the treatment of disorders associated with increased GBAP expression or activity, it is desirable to decrease the expression or activity of GBAP. In the treatment of disorders associated with decreased GBAP expression or activity, it is desirable to increase the expression or activity of GBAP.

Therefore, in one embodiment, GBAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GBAP. Examples of such disorders include, but are not limited to, an immune system disorder such as inflammation, actinic keratosis, acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, arteriosclerosis, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, bursitis, cholecystitis, cirrhosis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia, irritable

bowel syndrome, episodic lymphopenia with lymphocytotoxins, mixed connective tissue disease (MCTD), multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, primary thrombocythemia, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, trauma, and hematopoietic cancer including lymphoma, leukemia, and myeloma; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis, cancer of the breast, fibrocystic breast disease, and galactorrhea, a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; a nervous system disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; a cell signaling disorder including endocrine disorders such as disorders of the hypothalamus and pituitary resulting from lesions such as primary brain tumors, adenomas, infarction associated with pregnancy, hypophysectomy, aneurysms, vascular malformations, thrombosis, infections, immunological disorders, and complications due to head trauma; disorders associated with hyperpituitarism including acromegaly, giantism, and syndrome of inappropriate antidiuretic hormone (ADH) secretion (SIADH) often caused by benign adenoma; disorders associated with

hypothyroidism including goiter, myxedema, acute thyroiditis associated with bacterial infection; disorders associated with hyperparathyroidism including Conn disease (chronic hypercalemia); pancreatic disorders such as Type I or Type II diabetes mellitus and associated complications; disorders associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal cortex, hypertension associated with alkalosis; disorders associated with gonadal steroid hormones such as: in women, abnormal prolactin production, infertility, endometriosis, perturbations of the menstrual cycle, polycystic ovarian disease, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and, in post-menopausal women, osteoporosis; and, in men, Leydig cell deficiency, male climacteric phase, and germinal cell aplasia, hypergonadal disorders associated with Leydig cell tumors, androgen resistance associated with absence of androgen receptors, syndrome of 5 α-reductase, and gynecomastia; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing GBAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GBAP including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified GBAP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GBAP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of GBAP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GBAP including, but not limited to, those listed above.

In a further embodiment, an antagonist of GBAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of GBAP. Examples of such disorders include, but are not limited to, those immune system, reproductive, nervous system, and cell signaling disorders, and cell proliferative disorders including cancer, described above. In one aspect, an antibody which specifically binds GBAP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express GBAP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding GBAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of GBAP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of GBAP may be produced using methods which are generally known in the art. In particular, purified GBAP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind GBAP. Antibodies to GBAP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with GBAP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to GBAP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of GBAP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to GBAP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and

Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce GBAP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for GBAP may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between GBAP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering GBAP epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for GBAP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of GBAP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple GBAP epitopes, represents the average affinity, or avidity, of the antibodies for GBAP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular GBAP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from

about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the GBAP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of GBAP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of GBAP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al., supra.)

In another embodiment of the invention, the polynucleotides encoding GBAP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding GBAP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding GBAP. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding GBAP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked

inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and Somia, N. (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in GBAP expression or regulation causes disease, the expression of GBAP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in GBAP are treated by constructing mammalian expression vectors encoding GBAP and introducing these vectors by mechanical means into GBAP-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of GBAP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). GBAP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V.

and H.M. Blau, supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding GBAP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to GBAP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding GBAP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding GBAP to cells which have one or more genetic abnormalities with respect to the expression of GBAP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are

described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544; and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding GBAP to target cells which have one or more genetic abnormalities with respect to the expression of GBAP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing GBAP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res.169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding GBAP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotech. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full-length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for GBAP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of GBAP-coding RNAs and the synthesis of high levels of GBAP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic

replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of GBAP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding GBAP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding GBAP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding GBAP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased GBAP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding GBAP may be therapeutically useful, and in the treament of disorders associated with decreased GBAP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding GBAP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding GBAP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding GBAP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding GBAP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of

the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a pharmaceutical composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such pharmaceutical compositions may consist of GBAP, antibodies to GBAP, and mimetics, agonists, antagonists, or inhibitors of GBAP.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Pharmaceutical compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g.,

Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of pharmaceutical compositions may be prepared for direct intracellular delivery of macromolecules comprising GBAP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, GBAP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example GBAP or fragments thereof, antibodies of GBAP, and agonists, antagonists or inhibitors of GBAP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or

biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind GBAP may be used for the diagnosis of disorders characterized by expression of GBAP, or in assays to monitor patients being treated with GBAP or agonists, antagonists, or inhibitors of GBAP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for GBAP include methods which utilize the antibody and a label to detect GBAP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring GBAP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of GBAP expression. Normal or standard values for GBAP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibody to GBAP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of GBAP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding GBAP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of GBAP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of GBAP, and to monitor regulation of GBAP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding GBAP or closely related molecules may be used to identify nucleic acid sequences which encode GBAP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the

probe identifies only naturally occurring sequences encoding GBAP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the GBAP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:67-132 or from genomic sequences including promoters, enhancers, and introns of the GBAP gene.

Means for producing specific hybridization probes for DNAs encoding GBAP include the cloning of polynucleotide sequences encoding GBAP or GBAP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding GBAP may be used for the diagnosis of disorders associated with expression of GBAP. Examples of such disorders include, but are not limited to, an immune system disorder such as inflammation, actinic keratosis, acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, arteriosclerosis, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, bursitis, cholecystitis, cirrhosis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia, irritable bowel syndrome, episodic lymphopenia with lymphocytotoxins, mixed connective tissue disease (MCTD), multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, primary thrombocythemia, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, trauma, and hematopoietic cancer including lymphoma, leukemia, and myeloma; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis, cancer of the breast, fibrocystic breast disease, and galactorrhea, a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign

prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; a nervous system disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; a cell signaling disorder including endocrine disorders such as disorders of the hypothalamus and pituitary resulting from lesions such as primary brain tumors, adenomas, infarction associated with pregnancy, hypophysectomy, aneurysms, vascular malformations, thrombosis, infections, immunological disorders, and complications due to head trauma; disorders associated with hyperpituitarism including acromegaly, giantism, and syndrome of inappropriate antidiuretic hormone (ADH) secretion (SIADH) often caused by benign adenoma; disorders associated with hypothyroidism including goiter, myxedema, acute thyroiditis associated with bacterial infection; disorders associated with hyperparathyroidism including Conn disease (chronic hypercalemia); pancreatic disorders such as Type I or Type II diabetes mellitus and associated complications; disorders associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal cortex, hypertension associated with alkalosis; disorders associated with gonadal steroid hormones such as: in women, abnormal prolactin production, infertility, endometriosis, perturbations of the menstrual cycle, polycystic ovarian disease, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and, in post-menopausal women, osteoporosis; and, in men, Leydig cell deficiency, male climacteric phase, and germinal cell aplasia, hypergonadal disorders associated with Leydig cell tumors, androgen resistance associated with absence of androgen receptors, syndrome of 5 α-reductase, and gynecomastia; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal

hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding GBAP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered GBAP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding GBAP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding GBAP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding GBAP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of GBAP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding GBAP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development

of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding GBAP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding GBAP, or a fragment of a polynucleotide complementary to the polynucleotide encoding GBAP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding GBAP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding GBAP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of GBAP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid

quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described in Seilhamer, J.J. et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, incorporated herein by reference. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, antibodies specific for GBAP, or GBAP or fragments thereof may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity

(Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently

positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for GBAP to quantify the levels of GBAP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g.,

Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding GBAP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, e.g., Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding GBAP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may

also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, GBAP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between GBAP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with GBAP, or fragments thereof, and washed. Bound GBAP is then detected by methods well known in the art. Purified GBAP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding GBAP specifically compete with a test compound for binding GBAP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with GBAP.

In additional embodiments, the nucleotide sequences which encode GBAP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, in particular U.S. Ser. No. 60/144,595, U.S Ser. No. 60/150,460, and U.S. Ser. No. 60/159,849, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic

solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), pcDNA2.1 plasmid (Invitrogen, Carlsbad CA), or pINCY plasmid (Incyte Genomics, Palo Alto CA). Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using

PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (PE Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (PE Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VI.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programing, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and PFAM to acquire annotation using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full

length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full-length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:67-132. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

IV. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and

70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding GBAP occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation, trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

V. Chromosomal Mapping of GBAP Encoding Polynucleotides

The cDNA sequences which were used to assemble SEQ ID NO:67-132 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:67-132 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 5). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

The genetic map locations of SEQ ID NO:70, 74, 75, 77, 80, 86, 87, 90, 92, 93, 94, 97, 101, 106, 109, 111, 112, 113, 115, 117, 118, 121, and 128 are described in The Invention as ranges, or intervals, of human chromosomes. More than one map location is reported for SEQ ID NO:94, 101, 109, 111, and 115, indicating that previously mapped sequences having similarity, but not complete identity, to SEQ ID NO:94, 101, 109, 111, and 115 were assembled into their respective clusters. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters.

VI. Extension of GBAP Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:67-132 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this

fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and β-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems).

In like manner, the polynucleotide sequences of SEQ ID NO:67-132 are used to obtain 5' regulatory sequences using the procedure above, along with oligonucleotides designed for such extension, and an appropriate genomic library.

VII. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:67-132 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

VIII. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested

substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/µl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/µl RNase inhibitor, 500 µM dATP, 500 µM dGTP, 500 µM dTTP, 40 µM dCTP, 40 µM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60 °C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65 °C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60 °C. The arrays are washed for 10 min at 45 °C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45 °C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide

containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

IX. Complementary Polynucleotides

Sequences complementary to the GBAP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring GBAP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of GBAP. To inhibit transcription, a

complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the GBAP-encoding transcript.

X. Expression of GBAP

Expression and purification of GBAP is achieved using bacterial or virus-based expression systems. For expression of GBAP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express GBAP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of GBAP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding GBAP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, GBAP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from GBAP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified GBAP obtained by these methods can be used directly in the assays shown in Examples XI and XV.

XI. Demonstration of GBAP Activity

GTP-binding activity of GBAP is determined in an assay that measures the binding of GBAP

to α-P^{32}-labeled GTP. Purified GBAP is first blotted onto filters and rinsed in a suitable buffer. The filters are then incubated in buffer containing radiolabeled α-^{32}P-GTP. The filters are washed in buffer to remove unbound GTP and counted in a radioisotope counter. Non-specific binding is determined in an assay that contains a 100-fold excess of unlabeled GTP. The amount of specific binding is proportional to the activity of GBAP.

GTPase activity of GBAP is determined in an assay that measures the conversion of α-^{32}P-GTP to α-^{32}P-GDP. GBAP is incubated with α-^{32}P-GTP in buffer for an appropriate period of time, and the reaction is terminated by heating or acid precipitation followed by centrifugation. An aliquot of the supernatant is subjected to polyacrylamide gel electrophoresis (PAGE) to separate GDP and GTP together with unlabeled standards. The GDP spot is cut out and counted in a radioisotope counter. The amount of radioactivity recovered in GDP is proportional to GTPase activity of GBAP.

XII. Functional Assays

GBAP function is assessed by expressing the sequences encoding GBAP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT plasmid (Life Technologies) and pCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of GBAP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding GBAP and either CD64 or CD64-GFP. CD64 and CD64-

GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding GBAP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIII. Production of GBAP Specific Antibodies

GBAP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the GBAP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (PE Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-GBAP activity by, for example, binding the peptide or GBAP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XIV. Purification of Naturally Occurring GBAP Using Specific Antibodies

Naturally occurring or recombinant GBAP is substantially purified by immunoaffinity chromatography using antibodies specific for GBAP. An immunoaffinity column is constructed by covalently coupling anti-GBAP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing GBAP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of GBAP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/GBAP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and GBAP is collected.

XV. Identification of Molecules Which Interact with GBAP

GBAP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent.

(See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled GBAP, washed, and any wells with labeled GBAP complex are assayed. Data obtained using different concentrations of GBAP are used to calculate values for the number, affinity, and association of GBAP with the candidate molecules.

Alternatively, molecules interacting with GBAP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989, Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

GBAP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	67	1405545	LATRTUT02	1405545F6 (LATRTUT02), 1405545H1 (LATRTUT02), 2926327F7 (TLYMNOT04), 2926327T6 (TLYMNOT04)
2	68	1451265	PENITUT01	700515X14 (SYNORAT03), 758541H1 (BRAITUT02), 1348685F6 (PROSNOT11), 1451265H1 (PENITUT01), 1872777F6 (LEUKNOT02)
3	69	1556311	BLADTUT04	1556311H1 (BLADTUT04), 3221281T6 (COLNNON03), 3350311F6 (BRAITUT24), SBFA02256F1, SBFA01440F1, SBFA01098F1, SBFA04741F1
4	70	1901373	BLADTUT06	758057H1 (BRAITUT02), 1255886H1 (MENITUT03), 1887731X12C1 (BLADTUT07), 1901373H1 (BLADTUT06), 2866863H1 (KIDNNOT20), 3090943H1 (BRSTNOT19), 3215237H1 (TESTNOT07), 3719233H1 (PENCNOT10), 4319601H1 (BRADDIT02)
5	71	2367767	ADRENOT07	1331124F1 (PANCNOT07), 2367767H1 (ADRENOT07), 2367779F6 (ADRENOT07) 2782232F6 (BRSTNOT13), 3079286H2 (BRAIUNT01), 3584043T6 (293TF4T01), 4994696H1 (LIVRTUT11)
6	72	3090433	BRSTNOT19	312565H1 (LUNGNOT02), 841829R6 (PROSTUT05), 1340809H1 (COLNTUT03), 1842057H1 (COLNNOT07), 2693513F6 (LUNGNOT23), 3090433H1 (BRSTNOT19), 4895874H1 (LIVRTUT12)
7	73	3800591	SPLNNOT12	554715F1 (SCORNOT01), 882035X23 (THYRNOT02), 3042234F7 (BRSTNOT16), 3630695H1 (COLNNOT38), 3800591H1 (SPLNNOT12), 4975447H1 (HELATXT03)
8	74	5308471	MONOTXT02	790680R1 (PROSTUT03), 870507R1 (LUNGAST01), 948177R1 (PANCNOT05), 1682469T7 (PROSNOT15), 2897215H1 (KIDNTUT14), 5308471H1 (MONOTXT02)
9	75	5324322	FIBPFEN06	1001977R1 (BRSTNOT03), 1312045F1 (COLNFET02), 1334040F2 (COLNNOT13), 1488082F6 (UCMCL5T01), 1570077F1 (UTRSNOT05), 1929845H1 (COLNTUT03), 2306061H1 (NGANNOT01), 3127730F7 (LUNGTUT12), 3494367H1 (ADRETUT07), 3578924H1 (293TF3T01), 4619513H1 (ENDVNOT01), 4932823H1 (BRSTTUT20), 5324322H1 (FIBPFEN06)
10	76	067184	HUVESTB01	067184H1 (HUVESTB01), 067184R1 (HUVESTB01), 067184X12 (HUVESTB01), 067184X23C1 (HUVESTB01), 067184X29C1 (HUVESTB01), 968551H1 (BRSTNOT05), 2611874T6 (LUNGTUT10)
11	77	722896	SYNOOAT01	722896H1 (SYNOOAT01), 722896X19C1 (SYNOOAT01), 1433775T1 (BEPINON01), 1477633T6 (CORPNOT02), 2676923F6 (KIDNNOT19), 3230945H1 (COTRNOT01), 3389989H1 (LUNGTUT17)
12	78	1571739	UTRSNOT05	1571739H1 (UTRSNOT05), 1571739X12R1 (UTRSNOT05), 2799982H1 (PENCNOT01), 4059114F6 (BRAINOT21)

Table 1 (cont.)

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
13	79	1739479	HIPONON01	511157H1 (MPHGNOT03), 511157T6 (MPHGNOT03), 1739479H1 (HIPONON01), 2092446T6 (PANCNOT04), 3880948F6 (SPLNNOT11)
14	80	1999147	BRSTTUT03	1339243T6 (COLNTUT03), 1999147H1 (BRSTTUT03), 2094940X11F1 (BRAITUT02), 2670959T6 (ESOGTUT02), 3297709H1 (TLYJINT01), 3396927H1 (UTRSNOT16), SCBA00828V1, SCBA00615V1, SCBA04422V1, SCBA04646V1, SCBA01715V1, 5544151H1 (TESTNOC01)
15	81	2182085	SININOT01	767764R6 (LUNGNOT04), 1655010F6 (PROSTUT08), 1701703T6 (BLADTUT05), 1871360F6 (SKINBIT01), 2081835F6 (UTRSNOT08), 2411644H1 (BSTMNON02)
16	82	2216640	SINTFET03	489759H1 (HNT2AGT01), 2057454T6 (BEPINOT01), 2097739H1 (BRAITUT02), 2216640H1 (SINTFET03), 2325135H1 (OVARNOT02), 2361273R6 (LUNGFET05), 2667958H1 (ESOGTUT02), 3462348H1 (293TF2T01), 3478754H1 (OVARNOT11), 4163069F6 (BRSTNOT32)
17	83	2417361	HNT3AZT01	1394742F1 (THYRNOT03), 2417361F6 (HNT3AZT01), 2417361H1 (HNT3AZT01)
18	84	2454384	ENDANOT01	2454384H1 (ENDANOT01), 2454384T6 (ENDANOT01), 2589653T6 (LUNGNOT22), 2643485F6 (LUNGTUT08), 2723048H1 (LUNGTUT10), 3130367H1 (LUNGTUT12)
19	85	2610262	LUNGTUT08	1226946R6 (COLNNOT01), 1226946T6 (COLNNOT01), 2610262F6 (LUNGTUT08), 2610262H1 (LUNGTUT08)
20	86	2700075	OVARTUT10	604199R1 (BRSTTUT01), 1225126R1 (COLNTUT02), 1923323R6 (BRSTTUT01), 2301778R6 (BRSTNOT05), 2506882F6 (CONUTUT01), 2700075F6 (OVARTUT10), 2700075H1 (OVARTUT10), 2744960F6 (LUNGTUT11), 2833994F6 (TLYMNOT03), 2915413H1 (THYMFET03), 3647274H1 (ENDINOT01)
21	87	2786701	BRSTNOT13	754370R1 (BRAITUT02), 1426163R6 (BEPINON01), 1850667F6 (LUNGFET03), 1923562R6 (BRSTTUT01), 2215161F6 (SINTFET03), 2215161T6 (SINTFET03), 2498589H1 (ADRETUT05), 2991672F6 (KIDNFET02), 3028991H1 (HEARFET02), 3729514H1 (SMCCNON03), 5065467H1 (ARTFTDT01)
22	88	3068538	UTRSNOR01	908465R2 (COLNNOT09), 957130R6 (KIDNNOT05), 1301520F6 (BRSTNOT07), 1580628H1 (DUODNOT01), 2631247F6 (COLNTUT15), 3068538H1 (UTRSNOR01), 3532286T6 (KIDNNOT25)
23	89	5159072	BRSTTMT02	412241R1 (BRSTNOT01), 660435H1 (BRAINOT03), 881160H1 (THYRNOT02), 1304119F6 (PLACNOT02), 1324073F1 (LPARNOT02), 2520427H1 (BRAITUT21), 5159072H1 (BRSTTMT02)

Table 1 (cont.)

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
24	90	5519057	LIVRDIR01	066809H1 (HUVESTB01), 3279230H1 (STOMFET02), 5370305F6 (BRAINOT22), 5508943F6 (BRADDIR01), 5508943R6 (BRADDIR01), 5519057H1 (LIVRDIR01)
25	91	035379	HUVENOB01	035379H1 (HUVENOB01), 035379X11 (HUVENOB01), 035379X12 (HUVENOB01), 035379X13 (HUVENOB01), 035379X311D1 (HUVENOB01), 112161R1 (PITUNOT01), 1922877R6 (BRSTTUT01), 2133108F6 (ENDCNOT01), 3107232H1 (BRSTTUT15), 4798135H1 (LIVRTUT09), SCHA01519V1, g1802757
26	92	275354	TESTNOT03	275354H1 (TESTNOT03), 275354X1 (TESTNOT03), 166312T6 (BRSTNOT09), 2104284R6 (BRAITUT02), 2738788T6 (OVARNOT09), 3584082T6 (293TF4T01), SCGA07807V1
27	93	311658	LUNGNOT02	207452X12 (SPLNNOT02), 238306X85F1 (SINTNOT02), 26448 9H1 (HNT2AGT01), 311658H1 (LUNGNOT02), 1292829F6 (PGANNOT03), 1298271F1 (BRSTNOT07), 1488285H1 (UCMCL5T01), 2555757H1 (THYMNOT03), 2665984F6 (ADRENOT08), 2665984T6 (ADRENOT08), 3079209H1 (BRAIUNT01)
28	94	1251632	LUNGFET03	1251632H1 (LUNGFET03), 1251632X11 (LUNGFET03), 1251632X13 (LUNGFET03), 1316814T1 (BLADTUT02), 1384212F1 (BRAITUT08), 1711274F6 (PROSNOT16), 3128230H1 (LUNGTUT12), 4819602H1 (PROSTUT17), SZZZ00620R1
29	95	1331955	PANCNOT07	1363667X12 (LUNGNOT12), 1363667X13 (LUNGNOT12), SBBA01489F1, SBBA01528F1
30	96	1412614	BRAINOT12	1412614F6 (BRAINOT12), 1412614H1 (BRAINOT12), 2278130H1 (PROSNON01), 2278130T6 (PROSNON01), 5105388T6 (PROSTUS19)
31	97	1750781	LIVRTUT01	452712T6 (TLYMNOT02), 483862R6 (HNT2RAT01), 777729R6 (COLNNOT05), 1394724F1 (THYRNOT03), 1652134F6 (PROSTUT08), 1750781F6 (LIVRTUT01), 1750781H1 (LIVRTUT01), 1750781X305F1 (LIVRTUT01), 1750781X307D2 (LIVRTUT01), 3221477H1 (COLNNON03), SCHA02984V1, SXAA02156D1, SXAA00802D1
32	98	1821658	GBLATUT01	909674H1 (STOMNOT02), 1579095F1 (DUODNOT01), 1821658H1 (GBLATUT01), 1821658T6 (GBLATUT01), 2508922F6 (CONUTUT01), 2584263H1 (BRAITUT22), 5571821H1 (TLYMNOT08)
33	99	1872574	LEUKNOT02	305990F1 (HEARNOT01), 908252R2 (COLNNOT09), 1872574H1 (LEUKNOT02), 2051868F6 (LIVRFET02), 2285632R6 (BRAINON01), 3181732F6 (TLYJNOT01), 3285854F6 (HEAONOT05), 3332012H1 (BRAIFET01), SBWA02751V1, SBWA02849V1, SBWA04744V1, SBWA00180V1

Table 1 (cont.)

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
34	100	2590967	LUNGNOT22	1340471F6 (COLNTUT03), 2590967F6 (LUNGNOT22), 2771160F6 (COLANOT02), 3150287R6 (ADRENON04)
35	101	2824491	ADRETUT06	1381834X14 (BRAITUT08), 1381834X16 (BRAITUT08), 1381834X17 (BRAITUT08), 1381834X31 (BRAITUT08), 1972345F6 (UCMCL5T01), 2824491H1 (ADRETUT06), 3413970H1 (PTHYNOT04)
36	102	2825460	ADRETUT06	870873R6 (LUNGAST01), 1440326F1 (THYRNOT03), 2825460H1 (ADRETUT06), 2825460T6 (ADRETUT06), 4154518H1 (MUSLTMT01), 5068209H1 (PANCNOT23), SBLA03097F1
37	103	2871116	THYRNOT10	357664R6 (PROSNOT01), 1419595F1 (KIDNNOT09), 1419595T1 (KIDNNOT09), 1577877F6 (LNODNOT03), 1577877T1 (LNODNOT03), 2767635H1 (COLANOT02), 2871116F6 (THYRNOT10), 2871116H1 (THYRNOT10), 4650546H1 (PROSTUT20), SBHA03160F1, SBHA02613F1, SBHA02703F1
38	104	2942212	CONNTUT05	1270807H1 (TESTTUT02), 1270807X301D1 (TESTTUT02), 1270807X309D2 (TESTTUT02), 2942212H2 (CONNTUT05), g1924758
39	105	3685151	HEAANOT01	860843R1 (BRAITUT03), 1932207F6 (COLNNOT16), 1932207T6 (COLNNOT16), 2210580F6 (SINTFET03), 3043060H1 (HEAANOT01), 3685151H1 (HEAANOT01), 4960825H1 (TLYMNOT05)
40	106	4881515	UTRMTMT01	925415R1 (BRAINOT04), 1337450F6 (COLNNOT13), 1961288R6 (BRSTNOT04), 3581069H1 (293TF3T01), 3583842T6 (293TF4T01), 4881515H1 (UTRMTMT01), 5488514H1 (DRGTNON04), g1156606
41	107	5324681	FIBPFEN06	2455960T6 (ENDANOT01), 2458281F6 (ENDANOT01), 3834084F6 (PANCNOT17), 4046332H1 (LUNGNOT35), 5324681H1 (FIBPFEN06), g1733388, g1522074
42	108	5387651	BRAINOT19	810934T1 (LUNGNOT04), 822997R1 (KERANOT02), 1282647F1 (COLNNOT16), 1282647T1 (COLNNOT16), 1571430T6 (UTRSNOT05), 2208839F6 (SINTFET03), 2844787H1 (DRGLNOT01), 2908748H1 (THYMNOT05), 5387651H1 (BRAINOT19)
43	109	5595679	COLCDIT03	044292R6 (TBLYNOT01), 826501R1 (PROSNOT06), 1251632X12 (LUNGFET03), 1303934F1 (PLACNOT02), 1316814F1 (BLADTUT02), 1339567T1 (COLNTUT03), 2806159H1 (BLADTUT08), 2837021H1 (TLYMNOT03), 3037493H1 (BRSTNOT16), 3119883H1 (LUNGTUT13), 3395946H1 (LUNGNOT28), 3748742H1 (UTRSNOT18)
44	110	5782457	BRAXNOT03	532593R6 (BRAINOT03), 532593T6 (BRAINOT03), 5782457H1 (BRAXNOT03)

Table 1 (cont.)

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
45	111	760677	BRAITUT02	745006X13 (BRAITUT01), 760677H1 (BRAITUT02), 760677X19 (BRAITUT02), 763135X12 (BRAITUT02), 946075H1 (RATRNOT02), 953938H1 (SCORNON01)
46	112	1348567	PROSNOT11	1348567H1 (PROSNOT11), 1505075F6 (BRAITUT07), 1620627F6 (BRAITUT13), 2069105F6 (ISLTNOT01), 2417901F6 (HNT3AZT01), 2494683H1 (ADRETUT05), 3320166H1 (PROSBPT03)
47	113	1751354	LIVRTUT01	029909F1 (SPLNFET01), 029909R1 (SPLNFET01), 512371H1 (MPHGNOT03), 1439362F6 (PANCNOT08), 1751354F6 (LIVRTUT01), 1751354H1 (LIVRTUT01), 1900168F6 (BLADTUT06)
48	114	1976780	PANCTUT02	001347H1 (U937NOT01), 1755035X307D2 (LIVRTUT01), 1976780H1 (PANCTUT02), 2798389H1 (NPOLNOT01), 4050076H1 (SINTNOT18), 4228943H1 (BRAMDIT01), 4291877H1 (BRABDIR01), 5514957H1 (BRADDIR01), SCHA04173V1, SCHA02986V1, SCHA01162V1, SCIA02096V1
49	115	2048234	LIVRFET02	1553355F6 (BLADTUT04), 1929455F6 (COLNTUT03), 2048234H1 (LIVRFET02), 2699864T6 (OVARTUT10)
50	116	2111754	BRAITUT03	1335055F6 (COLNNOT13), 2105233R6 (BRAITUT03), 2111754H1 (BRAITUT03), 2111754R6 (BRAITUT03), 3706377H1 (PENCNOT07)
51	117	2123286	BRSTNOT07	411359F1 (BRSTNOT01), 411359R1 (BRSTNOT01), 708105R6 (SYNORAT04), 1322780F6 (BLADNOT04), 2123286H1 (BRSTNOT07), 2719651F6 (LUNGTUT10), 2880143F6 (UTRSTUT05), 3206153F6 (PENCNOT03), 3210501F6 (BLADNOT08), 3346625F6 (BRAITUT24), 3489118H1 (EPIGNOT01), 3605764H1 (LUNGNOT30), 4242993H1 (SYNWDIT01), 5089472H1 (UTRSTMR01)
52	118	2477507	SMCANOT01	488096H1 (HNT2AGT01), 1672690F6 (BLADNOT05), 1802830F6 (COLNNOT27), 1818538H1 (PROSNOT20), 2171841H1 (ENDCNOT03), 2477507H1 (SMCANOT01), 3434030F6 (PENCNOT05)
53	119	2759119	THP1AZS08	496782H1 (HNT2NOT01), 1251166H1 (LUNGFET03), 1289067F1 (BRAINOT11), 1295658T6 (PGANNOT03), 1510901F1 (LUNGNOT14), 1531583F1 (SPLNNOT04), 1533488F1 (SPLNNOT04), 1817447H1 (PROSNOT20), 2154846F6 (BRAINOT09), 2468875H1 (THYRNOT08), 2498852F6 (ADRETUT05), 2506652F6 (CONUTUT01), 2630812F6 (COLNTUT15), 2759119H1 (THP1AZS08), 2991227H1 (KIDNFET02), 3036646F6 (PENCNOT02), 3213032H1 (BLADNOT08)
54	120	2823818	ADRETUT06	618671R6 (PGANNOT01), 2823818H1 (ADRETUT06), 2950988F6 (KIDNFET01), g1679455

SUBSTITUTE SHEET (RULE 26)

Table 1 (cont.)

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
55	121	2859730	SININOT03	103901X6 (BMARNOT02), 510695H1 (MPHGNOT03), 1452088H1 (PENITUT01), 1527095F6 (UCMCL5T01), 2285371H1 (BRAINON01), 2843029H1 (DRGLNOT01), 2859730H1 (SININOT03)
56	122	2861155	SININOT03	875215T1 (LUNGAST01), 999673H1 (KIDNTUT01), 1425091R6 (BEPINON01), 2861155F6 (SININOT03), 2861155H1 (SININOT03), 2901915F6 (DRGCNOT01), 3621947H2 (ENDANOT03)
57	123	3002667	TLYMNOT06	227882F1 (PANCNOT01), 227882R1 (PANCNOT01), 260725H1 (HNT2RAT01), 1432542R1 (BEPINON01), 2474761F6 (SMCANOT01), 3002667H1 (TLYMNOT06), 3188977H1 (THYMNON04), 3461163H1 (293TF1T01), 4860339F6 (PROSTUT09)
58	124	3043734	HEAANOT01	3043734H1 (HEAANOT01), 3043734T6 (HEAANOT01), 3209823H1 (BLADNOT08), 5277071H1 (MUSLNOT01)
59	125	3294893	TLYJINT01	389234H1 (THYMNOT02), 1242886H1 (LUNGNOT03), 1539958T1 (SINTTUT01), 1870567H1 (SKINBIT01), 2069284F6 (ISLTNOT01), 2280217R6 (PROSNON01), 2353465T6 (LUNGNOT20), 2798990F6 (NPOLNOT01), 3180440H1 (TLYJNOT01), 3294893H1 (TLYJINT01), 3816962H1 (TONSNOT03), 5039889H2 (COLHTUT01), 5118831H1 (SMCBUNT01)
60	126	3349052	BRAITUT24	731775H1 (LUNGNOT03), 1449575H1 (PLACNOT02), 1899442F6 (BLADTUT06), 1967162T6 (BRSTNOT04), 2630025F6 (COLNTUT15), 2717821H1 (THYRNOT09), 3180478T6 (TLYJNOT01), 3349052H1 (BRAITUT24), 4523961F6 (HNT2TXT01), 5565623H1 (TLYMNOT08), 6141909H1 (BMARTXT03)
61	127	3357264	PROSTUT16	2378150F6 (ISLTNOT01), 2378150X304B1 (ISLTNOT01), 2378150X304D1 (ISLTNOT01), 2807493F6 (BLADTUT08), 2881251F6 (UTRSTUT05), 3357264F6 (PROSTUT16), 3357264H1 (PROSTUT16), 3593272H1 (293TF5T01), 4163652T6 (BRSTNOT32), 4821588F6 (PROSTUT17), 4872125H1 (COLDNOT01)
62	128	3576329	BRONNOT01	1444072F6 (THYRNOT03), 1649584T6 (PROSTUT09), 1720770X15C1 (BLADNOT06), 1720770X16C1 (BLADNOT06), 2204612F6 (SPLNFET02), 3576329H1 (BRONNOT01), SAFC01083F1
63	129	3805550	BLADTUT03	1416364F6 (BRAINOT12), 1553473H1 (BLADTUT04), 3232384H1 (COLNUCT03), 3287257H1 (HEAONOT05), 3539473H1 (SEMVNOT04), 3805550H1 (BLADTUT03)

Table 1 (cont.)

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
64	130	4546403	COLXTDT01	1687704F6 (PROSTUT10), 1962744R6 (BRSTNOT04), 2674742F6 (KIDNNOT19), 4546403H1 (COLXTDT01), 4632828T6 (GBLADIT02)
65	131	4767318	BRATNOT02	1345666R1 (BMARNOT02), 549352R1 (BEPINOT01), 1819757T6 (GBLATUT01), 2863295H1 (KIDNNOT20), 4767318H1 (BRATNOT02), SBLA03778F1, g3737930
66	132	4834527	BRAWNOT01	859906X38C1 (BRAITUT03), 1231225H1 (BRAITUT01), 1393681T6 (THYRNOT03), 1416996F6 (BRAINOT12), 2422475H1 (SCORNON02), 3999137R6 (HNT2AZS07), 4834527F6 (BRAWNOT01), 4834527H1 (BRAWNOT01), 5691642H1 (BRAUNOT02)

Table 2

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods & Databases
1	269	S59 T71 T146 T211 T73 S127 T133 S216	N12	GTP-binding protein: D79-M234, Y80-C239 ATP/GTP binding site (P-loop): G102-S109	GTP-binding protein; CgpA [Caulobacter crescentus] g3820578	BLAST-Genbank BLAST-DOMO MOTIFS
2	428	S59 S188 S200 S284 S367 S381 T399 T29 T193 T288 T354 S419		Beta transducin family, G-beta repeats: T269-L315 , F261-D293 L280-V294 , V185-V199 Signal peptide: M1-A35		ProfileScan MOTIFS BLIMPS-PRINTS HMMER-PFAM SPScan
3	562	S151 S152 T443 T444 S33 S104 S126 S127 S135 S216 S239 T350 T383 S450 T481 S146 T223 S287 S356 T434 T470 Y501	N125 N354 N445		Ras inhibitor [Homo sapiens] g190895	BLAST-Genbank
4	229	T108 S153 S9 S160 S215 T219 T142 S180	N111 N140 N198	ATP/GTP-binding site: G28-S35 Ras family: K23-T219 Ras transforming protein: V22-M43, A63-S85, P124-A137, L156-A178, D102-S145, K150-S180	Small GTP binding protein [Saccharomyces cerevisiae] g1171484	BLAST-Genbank MOTIFS HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS BLAST-DOMO
5	360	T108 S360 S115 T217 T264 S295 S296 S35 S52 S160 S174 T206 T249	N149 N287 N327 N351	WD domain, G-beta repeats: M1-T64, M27-K41, F274-K306	Similar to WD domain, G-beta repeat protein [C. elegans] g3880929	BLAST-Genbank HMMER-PFAM ProfileScan BLIMPS-PRINTS
6	460	T18 T107 T123 S149 S199 S280 S336 S369 S71 T106 S387 Y302 Y400	N270 N350	Signal peptide: M1-A57	Rabin3 [Rattus norvegicus] g624225	BLAST-Genbank SPScan

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods & Databases
7	239	S234 S25 T47 T52 S98 T190 T206 S236 S223	N188	Phosducin: L20-I179, S25-I179, E30-D239	Phosducin-like protein [Homo sapiens] g4104075	BLAST-Genbank BLAST-PRODOM BLAST-DOMO
8	334	T225 T235 S260 T4 S45 S63 S133 S162 S193 T279 T308		ATP/GTP-binding site (P-loop): G150-S157 GTP1/OBG family: L75-D89, I146-Q166 G-protein, alpha subunit: I79-L87	GTP-binding protein homolog [L. braziliensis] g2570231	BLAST-Genbank MOTIFS BLIMPS-BLOCKS BLIMPS-PRINTS
9	341	S91 T122 S185 T199 T228 S65 T85 S323		Signal peptide: M1-A61 WD domain,G-beta repeats: L164-D196, C173-P217, V183-L197, S185-W195	Putative WD-40 repeat protein [Arabidopsis thaliana] g4191773	SPScan BLAST-Genbank MOTIFS ProfileScan HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS
10	513	T29 T72 T109 S124 S136 S215 T341 T481 T501 S65 T245 T330 S338 T372 T386 S437 S451 T473 Y228 Y254	N242 N417	Beta-transducin family, G-beta repeats: F345-N377, K210-N242, E303-G335, S366-W376, N353-V400, L229-F243, I364-M378	Similar to WD domain G-beta repeats protein [C. elegans] g3875246	BLAST-Genbank MOTIFS HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS ProfileScan
11	186	T61 S80 S107 S163 S31 T66 S183	N64 N148	ARF-family: N6-S186, P51-S90, M95-L149 GTP-binding, SAR1 protein: F78-K103, I123-I144 ATP/GTP binding site (P-loop): G27-T34	Similar to ADP-ribosylation factor [C. elegans] g3881189	BLAST-Genbank HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS MOTIFS

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods & Databases
12	204	S184 S203 S34 S152 T14 T20 T25 T62 S86		Ras family: K5-M189 Ras transforming protein: M1-E150, V4-T25, V113-L126 ATP/GTP binding site (P-loop): G10-S17	Ras-like protein, rit [Mus musculus] g1656005	BLAST-Genbank HMMER-PFAM BLIMPS-PRINTS BLAST-DOMO MOTIFS
13	100	S31 S46 T52 T61 S84 S4 S26 S27 T86		Beta-transducin, WD repeats: L81-M95, V70-S100, M1-S100	Similar to beta-transducin [C. elegans] g3875373; Alzheimer's disease protein [Homo sapiens] GeneSeq W21578	BLAST-Genbank MOTIFS BLIMPS-BLOCKS ProfileScan BLIMPS-PRINTS BLAST-PRODOM
14	795	T569 S776 S54 S188 S201 T248 T249 T298 S306 S368 T422 S466 T561 S586 S625 S678 T731 S777 S13 T42 S120 T134 T174 S213 S254 T266 S391 S415 S588 S620 S694 T742	N52 N421 N559 N585 N708	WD domain, G-beta repeats: L108-L139, L147-K179, T168-W178, Y227-K259, L126-N140, M166-A180	Phospholipase A2-activating protein [Rattus Norvegicus] g1017706	BLAST-Genbank BLAST-PRODOM BLAST-DOMO HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS
15	393	S48 S61 T143 T334 T148 T200 S208 T212 T245 S266 S325	N182 N197	WD domain, G-beta repeats: L121-A153, L357-R389, P322-F369, L140-S154	Putative WD-repeat protein [Arabidopsis thaliana] g4263521	BLAST-Genbank HMMER-PFAM ProfileScan BLIMPS-PRINTS
16	485	S31 S108 S222 S321 S346 S357 T84 T125 T137 T151 T187 S227 T268 S395 T403 S409 T437 Y92 Y261		Beta-transducin, WD repeats: L129-L143, V219-T233, S262-W272, V387-G401, L429-V443, L452-G468	Notchless protein [Xenopus laevis] g3687833	BLAST-Genbank MOTIFS HMMER-PFAM ProfileScan BLIMPS-BLOCKS BLIMPS-PRINTS BLAST-DOMO BLAST-PRODOM

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods & Databases
17	199	T32 T91 S177 T56 S153 S186 Y149		ATP/GTP-binding site (P-loop): G15-T22 Transforming protein, p21: L9-H30, T32-K48, I50-S72, Q115-L128, Y149-A171 Ras protein: K5-E151	Rab7 [Mus musculus] g1050551	BLAST-Genbank MOTIFS BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO
18	163	T18 T46 S120 S5 T151 T83 S125	N81 N159		Rhotekin [Mus musculus] g1293145	BLAST-Genbank
19	290	S56 S84 T234 S41 T91 T132 T234 T11 T47 T80 T194	N89 N188	Beta-transducin, WD-repeats: S41-W51, F195-D227, L238-N270, L214-I228, L257-M271, T203-S249	Similar to beta-transducin; [C. elegans] g3875373; Alzheimer's disease protein [Homo sapiens] GeneSeq W21578	BLAST-Genbank MOTIFS HMMER-PFAM BLIMPS-BLOCKS ProfileScan BLIMPS-PRINTS BLAST-PRODOM
20	705	T277 T364 S393 S448 S479 S483 T554 T568 S586 S239 S250 T374 S379 T398 S485 T528	N274	Beta-transducin, WD-repeats: L390-L404, L370-D403, L413-R445	Similar to WD domain G-beta repeat prot. [C. elegans] g3880340; 70kD tumor-specific antigen [R. norvegicus] g2505957	BLAST-Genbank HMMER-PFAM BLAST-PRODOM BLAST-DOMO BLIMPS-BLOCKS BLIMPS-PRINTS MOTIFS
21	454	T426 S451 S28 S51 T81 T89 T166 S214 T241 S264 T305 S343 S185 T193 S421	N58	ATP/GTP-binding site (P-loop): G73-S80 Cell division control protein: V47-P240	Similar to Drosophila melanogaster septin (sep2) [Homo sapiens] g1503988	BLAST-Genbank BLAST-PRODOM BLAST-DOMO MOTIFS
22	433	S169 T239 T292 S309 S382 S129 S297 Y60 Y101 Y315	N338	Protein GTPase activating protein: L8-S169 PH domain: Y138-Q355, Q191-I351, P210-E375	RhoGAP protein [Homo sapiens] g312212	BLAST-Genbank BLAST-PRODOM BLAST-DOMO

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods & Databases
23	406	T83 S143 S303 T75 T115 T126 T211 S216 T289 T315 Y247	N184 N401 N402		Rab 9 effector, P40 [Homo sapiens] g2217970	BLAST-Genbank
24	229	S7 S127 T50 S178		ATP/GTP-binding site (P-loop): G40-T47 Ras family: K35-L217 Transforming protein, p21: F34-A55, R57-R73, V75-K97, N139-L152	Rab GTPase, Rab33B [Mus musculus] g2516239	BLAST-Genbank MOTIFS HMMER-PFAM BLIMPS-PRINTS BLAST-DOMO
25	670	T28 T45 S69 S3 S108 T277 S406 S6 T52 T82 S91 S102 S126 S609 S158 S197 T213 S217 T281 S323 S416 T419 T428 T474 S496 T540 S624 T664	N343	G-beta WD repeat domain: F386-D424, L411-T425, Y429-D465, L469-D504, L510-D545, L549-D585, K589-S629, M633-T669 Beta-transducin Trp-Asp repeats signature: C401-I447	Beta transducin-like protein [Podospora anserina] g607003	BLAST-GenBank BLIMPS-BLOCKS BLIMPS-PRINTS HMMER-PFAM MOTIFS ProfileScan
26	445	T17 T48 T126 T160 T293 T364 T97 T132 S201 S217 S305 T322 S357 S434 Y339	N46 N95 N355	G-beta WD repeat domain: L62-N95, V82-L96, F124-M138, F297-V311 Beta-transducin Trp-Asp repeats signature: S316-A356 SOF1 protein, WD repeat: D129-V277, F309-V444	Beta-transducin [Schizosaccharomyces pombe] g3393019	BLAST-GenBank BLAST-DOMO BLAST-PRODOM BLIMPS-BLOCKS BLIMPS-PRINTS HMMER-PFAM MOTIFS
27	236	S24 S60 S86 T181 S117 S140		GYP7, GTPase activating protein: M1-I155	GTPase activating protein [Yarrowia lipolytica] g2370595	BLAST-GenBank BLAST-PRODOM MOTIFS

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods & Databases
28	498	S97 T158 S247 S281 S425 S468 S494 T84 S176 T355 T474 Y239		G-beta WD repeat domain: L188-Q220, L446-G479, M466-P480 Beta-transducin Trp-Asp repeats signature: F200-A245	similarity to guanine nucleotide binding protein [Caenorhabditis elegans] g3878300	BLAST-GenBank BLIMPS-BLOCKS BLIMPS-PRINTS HMMER-PFAM MOTIFS ProfileScan
29	334	S63 S104 S148 S189 T208 S276 S50 T110 S118 T124 S152 T160 T237 T326	N265	G-beta WD repeat domain: L41-G73, I83-D115, L102-V116, L125-D157, L167-D199, I210-D242 Beta-transducin Trp-Asp repeats signature: S49-A308 Signal peptide: M1-A47	similar to guanine nucleotide binding protein [Caenorhabditis elegans] g3874290	BLAST-GenBank BLAST-PRODOM BLIMPS-BLOCKS BLIMPS-PRINTS HMMER-PFAM MOTIFS ProfileScan SPScan
30	292	S102 T145 S188 S52 T89 S204 S222 S283	N209	Protein with WD repeat: P7-W129 Signal peptide: M1-S68	F-box protein FBX16 [Mus musculus] g6456114	BLAST-PRODOM BLAST-GenBank MOTIFS SPScan
31	588	T184 T76 T137 S139 T161 T174 T183 S285 T351 T375 S432 T473 S488 S213 T265 S389 S394 T412 T546	N159	G-beta WD repeat domain: A293-E331, C337-T375, Y379-D417, I404-L418, E460-D497, T506-S543, G547-A586 Beta-transducin Trp-Asp repeats signature: A308-E354, L393-Q441	TipD (sequence similarity to Beta-transducin family) [Dictyostelium discoideum] g2407788	BLAST-GenBank BLAST-PRODOM BLIMPS-BLOCKS BLIMPS-PRINTS HMMER-PFAM MOTIFS ProfileScan
32	326	T50 T84 S98 S142 T261 T65 T148 T178 T189 T221	N187	G-beta WD repeat domain: L120-N153, I140-L154		BLIMPS-BLOCKS BLIMPS-PRINTS HMMER-PFAM MOTIFS
33	453	T157 T218 T248 S320 S347 S412 S7 T236 S290 T396 T406 Y63	N59 N225	G-beta WD repeat domain: D180-E211, A198-V212		BLIMPS-BLOCKS BLIMPS-PRINTS HMMER-PFAM MOTIFS

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods & Databases
34	161	T137 T18 T102 Y96		DMR-N9 protein: K93-S148	DMR-N9 (homology to WD repeat sequences) [Mus musculus] g817954	BLAST-GenBank BLAST-PRODOM MOTIFS
35	684	T173 S25 S43 S74 S83 S127 S152 S154 S182 T316 T331 T341 S372 T535 T606 S623 T138 T151 S168 S238 S299 T336 T422 S476 T506 T530 T628 T647	N526 N621	ATP/GTP-binding site motif A (P-loop): G267 Elongation factor 1 alpha protein (GTP-binding) domain: D485-E684 Elongation factor Tu domain: K258-D658, N262-K273, M343-G374, R664-G677 GTP-binding elongation factors signature: A249-E420, N262-T275, K294-P346, T341-F351, T357-V368, L401-Q410, P443-I682 RAS transforming protein: K258-V439	eRFS (related to eukaryotic release factor 3) [Mus musculus] g4566435	BLAST-GenBank BLAST-DOMO BLAST-PRODOM BLIMPS-BLOCKS BLIMPS-PRINTS HMMER-PFAM MOTIFS ProfileScan
36	366	S342 T52 S71 T102 T119 T224 T324 T66 S195 S271 T353 Y225	N32	G-beta WD repeat domain: V146-L160, L284-I298 Signal Peptide: M1-T56		BLIMPS-PRINTS MOTIFS SPScan
37	339	S152 S183 T107 T115		Beta-transducin Trp-Asp repeats signature: N101-L162 Trp-Asp repeats-containing protein: R54-A172 Transmembrane domain: A300-I323	Hypothetical trp-asp repeats containing protein [Schizosaccharomyces pombe] g3850059	BLAST-GenBank BLAST-DOMO BLAST-PRODOM HMMER MOTIFS

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods & Databases
38	213	T29 T134 S153 T181 S200 T92 T129 S207		ATP/GTP-binding site motif A (P-loop): G15 GTP-binding protein signature (Arf1, Ran): W5-E179 Ras family signature: R10-C213 Transforming protein p21: F9-E30, R32-R48, E51-S73, Y114-L127, Y149-I171 signal peptide: M1-V19	Rab-related GTP-binding protein [Homo sapiens] g1491714	BLAST-GenBank BLAST-DOMO BLAST-PRODOM BLIMPS-PRINTS HMMER-PFAM MOTIFS SPScan
39	393	S209 T363 S60 S99 S119 S135 T144 T147 S174 S210 T350 S359 S370 T371		G-beta WD repeat domain: G33-D69, K73-D110, L97-A111, W114-N152, L236-K276, I263-L277 Signal peptide: M1-T43	similar to beta-transducin [Caenorhabditis elegans] g860695	BLAST-GenBank BLIMPS-BLOCKS BLIMPS-PRINTS HMMER-PFAM MOTIFS SPScan
40	399	S86 T191 S219 S224 S254 S275 S308 S59 S72 T96 S373 S385 T394	N88 N106 N321 N322	ATP/GTP-binding site motif A (P-loop): G68 G-protein alpha subunit: R63-Q78 GTP-binding protein GTR1: A57-D294 Ras transforming protein: K61-L203	Gtr2 homolog, novel small GTPase subfamily [Schizosaccharomyces pombe] g3560242	BLAST-GenBank BLAST-DOMO BLAST-PRODOM BLIMPS-PRINTS MOTIFS
41	412	T106 S337 S391 S29 S30 S41 S130 S154 S207 S231 S326 S82 S97 T212 S220	N367	G-beta WD repeat domain: C184-E217, L204-Y218 signal peptide: M1-G18	Putative transcriptional regulation protein, trp-asp repeat containing [Schizosaccharomyces pombe] g3766375	BLAST-GenBank BLIMPS-BLOCKS BLIMPS-PRINTS HMMER-PFAM MOTIFS SPScan

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods & Databases
42	163	S15 S17 S71 T114 Y49			Arf-like 2 binding protein BART1 [Homo sapiens] g4426962	BLAST-GenBank MOTIFS
43	514	S113 T174 S263 S297 S441 S484 S510 T100 S192 T371 T490 Y255		G-beta WD repeat domain: L204-Q236, L462-G495, M482-P496 Beta-transducin Trp-Asp repeats signature: F216-A261	similarity to guanine nucleotide binding protein [Caenorhabditis elegans] g3878300	BLAST-GenBank BLIMPS-BLOCKS BLIMPS-PRINTS HMMER-PFAM MOTIFS ProfileScan
44	67	T30 S15 Y18		G-protein gamma subunit: E2-L67, M9-R24, K10-P57, D45-G62 Prenyl group binding site (CAAX box): V64	G gamma protein [Mus musculus] g7259257	BLAST-GenBank BLIMPS-BLOCKS BLIMPS-PRINTS HMMER-PFAM MOTIFS
45	315	T148 S162 S209 S244 S252 S45 T48 S132 S140 S158 T214 S244	N79	WD40 domains/G-beta repeats: Q15-N53, G57-N95, G99-D137, P143-D179, G223-D263 WD/G-beta profiles: L71-Q116, T114-V161 WD/G-beta repeat signature: V250-L264	Contains similarity to G beta repeats (PROSITE:PS00670) of the beta-transducin family [Caenorhabditis elegans] g1086900	BLAST-GenBank MOTIFS ProfileScan HMMER-PFAM
46	504	T268 T99 T193 S323 S324 T409 T493 T91 T98 T133 T185 T234 T259 T264 T287 T337 S415 S498	N37 N295	WD40 domains/G-beta repeats: A211-D250, E254-S292, A296-A331, G338-D378, R382-D420 WD/G-beta profiles: T396-I442, T268-A316, C355-F400 WD/G-beta signatures: L407-L421, V279-V293 WD repeat protein-like region: I4-A226	Similar to S. cerevisiae PRP19 protein; simliar to G-beta repeat region of guanine nucleotide binding protein [Caenorhabditis elegans] g727450	BLAST-GenBank BLAST-PRODOM MOTIFS BLIMPS-PRINTS ProfileScan HMMER-PFAM

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods & Databases
47	522	S84 S315 S510 T20 S50 S57 S74 S116 S122 S128 S161 S185 T274 T300 S339 S345 S357 S367 T373 S459 T474 S136 S143 T174 S200 T300 S315 S356 S385 S420 T492	N226 N355		SAPK (stress activated protein kinase) interacting protein (similar to ras inhibitor) [Gallus gallus] g4929812	BLAST-GenBank MOTIFS
48	316	T109 S27 S86 S188 S7 S8 S82 T96 T105	N29 N136 N186	Pleckstrin homology (PH) domains: S3-N45, I59-Q301 RhoGAP domain: P140-N291 GTPase protein-like region: G125-L307	Beta2-chimaerin [Homo sapiens] g457230	BLAST-GenBank BLAST-PRODOM BLAST-DOMO HMMER-PFAM MOTIFS BLIMPS-PRINTS BLIMPS-PRODOM
49	387	S97 S199 T249 S342 S369 S382 T54 T182 T381		ATP/GTP-binding site motif (P-loop): G155-S162 GTP1/OBG GTP-binding protein family signatures: V151-A171, K172-I190, V200-G215, G217-D235 GTP-binding protein-like region: F15-P173 RAS transforming protein-like region: L145-L296	GTP-binding protein [Aquifex aeolicus] g2984292	BLAST-GenBank BLAST-PRODOM BLAST-DOMO BLIMPS-BLOCKS BLIMPS-PRINTS MOTIFS

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods & Databases
50	334	T228 T308 S65 S91 T224 T228 T262 S34 S81 T224 T262 S286 T324	N108 N257 N322	ATP/GTP-binding site motif (P-loop): G149-S156 Ras domain: R144-M334 p21/ras-related transforming protein signatures: Y143-S164, N166-L182, H248-D261, F282-K304 Ras transforming protein-like region: I140-E284	NOEY2 putative tumor suppressor [Homo sapiens] g4100355	BLAST-GenBank BLAST-PRODOM BLAST-DOMO HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS MOTIFS
51	551	T199 S38 T62 S85 T116 S169 S351 T379 S421 S422 S456 S12 S22 S150 T366 S383 T482 Y404 Y449	N133 N148 N179 N293 N296	Regulator of chromosome condensation (RCC1)/ guanine nucleotide dissociation stimulator domains: E117-S169, D170-D222, T223-D274, E275-G292, G328-G339 RCC1 signatures: V157-L167, V262-L272	UVB-resistance protein UVR8 [Arabidopsis thaliana] g5478530	BLAST-GenBank BLAST-PRODOM HMMER-PFAM PROFILESCAN BLIMPS-PRINTS MOTIFS
52	308	S152 T230 S266 S299 S19 S22 S240	N76	WD40 domains/G-beta repeats: Q33-R73, W79-T119, W126-K181, W188-T230, P241-K276, S11-A50 Sec13 related/WD repeat protein-like region: R73-I177 WD/G-beta profile: G11-A50	Sec13-related protein [Arabidopsis thaliana] g3150415	BLAST-GenBank HMMER-PFAM PROFILESCAN BLIMPS-PRINTS BLAST-PRODOM MOTIFS

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods & Databases
53	949	S206 S514 T22 S216 T226 S273 T315 S663 T745 T908 T155 S232 S258 T350 S359 S472 S609 S776 S837 S913 Y682 Y862	N114	WD40 domains/G-beta repeats: V199-K237, V248-S284, G287-H326 Drosophila lethal(2) giant larvae tumor suppressor protein signature: K221-P244, A353-E377		HMMER-PFAM BLIMPS-PRINTS MOTIFS
54	227	S11 T113 S173 T155 S173	N38	ATP/GTP-binding site motif (P-loop): G37-T44 Ras family domain: K32-C227 p21/ras-related transforming protein signatures: F31-D52, S54-K70, V72-T94, D134-M147, F169-I191 Ras transforming protein-like region: F27-T172	GTP-binding protein [Bos taurus] g162764	BLAST-GenBank HHMER-PFAM BLIMPS-PRINTS BLAST-DOMO BLAST-PRODOM MOTIFS
55	474	T430 S98 S118 S309 S450 S463 T66 S130 T141 S241 S289 S309 S389 S450	N179 N185	WD40 domains/G-beta repeats: D70-Q109, T120-N159, E164-D202 G-beta repeat signature: L146-V160 WD repeat/coronin protein-like region: I208-Q467	Coronin-2 [Mus musculus] g4895039	BLAST-GenBank HMMER-PFAM BLAST-PRODOM BLAST-DOMO MOTIFS
56	547	S16 T77 S85 S90 S112 S114 T132 S160 T166 T225 S248 S438 S491 S526 S125 S267 T299 T305 S504	N101 N110 N147 N297	WD40 domains/G-beta repeats: G159-N197, C312-A353, G357-D396 WD40/G-beta signatures: V245-A259, L428-T442	Guanine nucleotide-binding protein beta 5 [Mesocricetus auratus] g1001939	BLAST-GenBank HMMER-PFAM BLIMPS-PRINTS MOTIFS

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods & Databases
57	686	T331 S431 T637 S34 S169 S554 S28 S124 S192 S273 S341 T366 S426 S449 S470	N26 N44 N271 N424 N628	G-beta profile: S106-S152	Beta-transducin-like protein [Podospora anserina] g607003	BLAST-GenBank PROFILESCAN HMMER-PFAM
58	93	S15 T2 S3 T24			HP protein (RhoGAP ortholog) [Homo sapiens] g2559002	BLAST-GenBank MOTIFS
59	521	S63 S223 T64 T117 S147 S159 S195 S200 T214 S271 S401 S448 T49 S110 S195 T235 T280 T439	N71 N108 N381	Amino acyl tRNA ligase motif: P173-T183	GTPase activating protein [Schizosaccharomyces pombe] g3150248	BLAST-GenBank MOTIFS
60	751	T287 S543 T61 S275 S345 T430 T474 T565 T676 S705 S726 T727 S57 T63 T70 T287 S345 T389 T432 S458 T479 T518 T538	N344 N640	GTP binding elongation factor Tu family domain: E44-T530 Elongation factor G C-terminus domain: L556-T727 GTP binding elongation factor signatures: N48-T61, Q97-A105, N117-F127, R133-V144, F169-R178	Elongation factor G [Rattus norvegicus] g310102	BLAST-GenBank HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS ProfileScan BLAST-PRODOM BLAST-DOMO MOTIFS
61	666	T492 S615 S619 T35 S142 T177 T212 S224 S270 T353 S403 T456 T471 T500 T550 S560 S572 T378 S403 S496 T509 T608 T611 T625	N75 N582		Rho target rhophilin [Mus musculus] g1176422	BLAST-GenBank MOTIFS

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods & Databases
62	746	S22 T98 S571 T46 S53 S61 S66 S70 S71 T97 S14 S126 S127 T165 T184 T190 S249 S279 S323 S430 S519 S680 S736 S115 T190 T237 S349 S436 T444 S567 S598 S601 T613 S652 T741		WD40 domains/G-beta repeats: T403-E441, R570-H606, Q610-D648, T653-H691, L704-T746, C418-A461 G-beta repeat signature: L428-V442 Trp-Asp repeat protein-like region: S22-L407	Bop1 growth control protein [Mus musculus] g1679772	BLAST-GenBank BLAST-PRODOM BLAST-DOMO MOTIFS BLIMPS-PRINTS ProfileScan HMMER-PFAM
63	212	S105 S142 S148 S162 S167 S44 T56 T101 S162 S190	N131	ATP/GTP-binding site motif (P-loop): G25-T32 Ras family domain: K20-C212 ADP-ribosylation factor family domain: P6-R183 p21/ras-related transforming protein signatures: F19-T40, A42-K58, L60-T82, S122-L135, A158-L180 Ras transforming protein-like region: Y15-I155	Rab19 [Mus musculus] g2598565	BLAST-GenBank HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS BLAST-DOMO BLAST-PRODOM MOTIFS

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods & Databases
64	307	T275 S276 T15 S25 T99 S164 S201 S6 S270 T293	N196 N291	WD40 domains/G-beta repeats: M1-I49, L60-D98, E102-Q140 Sterile alpha motif (SAM): E161-R225 WD/G-beta signatures: L36-V50, L127-F141 G-beta profile: L74-P122	Hypothetical trp-asp repeats protein [C. elegans] SwissProt Q93847	BLAST-SwissProt HMMER-PFAM BLIMPS-PRINTS ProfileScan MOTIFS
65	378	S137 T167 T193 S202 S237 S276 S290 S310 S362 S82 T150 T158 T199 S362 T368		WD40 domains/G-beta repeats: H72-L110, L116-D155, L241-D279 G-beta profiles: S137-C175, S87-C133, I255-S312	WD repeat protein [Schizosaccharomyces pombe] g5701965	BLAST-GenBank HMMER-PFAM ProfileScan MOTIFS
66	466	S6 T24 S69 T209 S246 S357 T450 S181 S236 S242 T322 T407 T450	N448	RasGEF domain: V197-E397 Guanine nucleotide releasing protein-like region: P201-S432	Putative guanine-nucleotide releasing factor [Drosophila affinis] g2981229	BLAST-GenBank HMMER-PFAM BLAST-PRODOM BLAST-DOMO

Table 3

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
67	434-478	Cardiovascular (0.238) Reproductive (0.238) Hematopoietic/Immune (0.190)	Cancer (0.429) Inflammation/Trauma (0.524) Cell Proliferation (0.095)	pINCY
68	380-424 551-595	Nervous (0.185) Reproductive (0.167) Gastrointestinal (0.148)	Cancer (0.444) Cell Proliferation (0.315) Inflammation/Trauma (0.278)	pINCY
69	433-477	Reproductive (0.429) Nervous (0.142) Hematopoietic/Immune (0.142)	Cancer (0.714) Inflammation/Trauma (0.142)	pINCY
70	684-728	Reproductive (0.333) Nervous (0.178) Cardiovascular (0.111)	Cancer (0.467) Cell Proliferation (0.244) Inflammation/Trauma (0.267)	pINCY
71	219-263	Hematopoietic/Immune (0.257) Reproductive (0.229) Gastrointestinal (0.143)	Cell Proliferation (0.400) Inflammation/Trauma (0.429) Cancer (0.314)	pINCY
72	865-912	Gastrointestinal (0.286) Reproductive (0.286) Cardiovascular (0.238)	Cancer (0.667) Cell Proliferation (0.143) Inflammation/Trauma (0.238)	pINCY
73	900-944	Reproductive (0.229) Hematopoietic/Immune (0.157) Nervous (0.157)	Cancer (0.422) Inflammation/Trauma (0.349) Cell Proliferation (0.205)	pINCY
74	109-153 919-963	Reproductive (0.270) Gastrointestinal (0.162) Cardiovascular (0.135)	Cancer (0.405) Cell Proliferation (0.270) Inflammation/Trauma (0.324)	pINCY
75	1352-1396 1568-1612	Reproductive (0.296) Gastrointestinal (0.167) Nervous (0.167)	Cancer (0.509) Inflammation/Trauma (0.269) Cell Proliferation (0.157)	pINCY
76	541-585 1189-1233	Reproductive (0.238) Cardiovascular (0.190) Gastrointestinal (0.190)	Cancer (0.524) Inflammation/Trauma (0.310) Cell Proliferation (0.143)	PBLUESCRIPT
77	110-154	Reproductive (0.250) Nervous (0.224) Hematopoietic/Immune (0.132) Gastrointestinal (0.132)	Cancer (0.355) Inflammation/Trauma (0.342) Cell Proliferation (0.211)	PSPORT1
78	218-262	Reproductive (0.375) Nervous (0.188) Urologic (0.188)	Cancer (0.562) Inflammation/Trauma (0.250)	pINCY

Table 3 (cont.)

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
79	380-424	Hematopoietic/Immune (0.227) Nervous (0.227) Reproductive (0.227)	Inflammation/Trauma (0.636) Cancer (0.364)	PSPORT1
80	217-261	Reproductive (0.275) Gastrointestinal (0.196) Nervous (0.196)	Cancer (0.431) Inflammation/Trauma (0.451) Cell Proliferation (0.196)	PSPORT1
81	488-532 812-856	Reproductive (0.301) Nervous (0.151) Gastrointestinal (0.130)	Cancer (0.466) Inflammation/Trauma (0.288) Cell Proliferation (0.151)	pINCY
82	595-639	Reproductive (0.333) Developmental (0.148) Gastrointestinal (0.148)	Cancer (0.444) Cell Proliferation (0.370) Inflammation/Trauma (0.333)	pINCY
83	219-263	Hematopoietic/Immune (0.400) Gastrointestinal (0.200) Cardiovascular (0.100)	Inflammation/Trauma (0.429) Cell Proliferation (0.357) Cancer (0.286)	pINCY
84	164-208	Cardiovascular (0.667) Nervous (0.222) Hematopoietic/Immune (0.111)	Cancer (0.556) Cell Proliferation (0.111)	PBLUESCRIPT
85	487-531 757-801	Reproductive (0.182) Cardiovascular (0.091)	Cancer (0.308) Cell Proliferation (0.231) Inflammation/Trauma (0.154)	pINCY
86	325-369 811-855	Hematopoietic/Immune (0.288) Reproductive (0.197) Cardiovascular (0.136)	Inflammation (0.394) Cancer (0.318) Cell Proliferation (0.212)	pINCY
87	163-207	Reproductive (0.218) Nervous (0.172) Gastrointestinal (0.138)	Cancer (0.448) Cell Proliferation (0.218) Inflammation (0.207)	pINCY
88	362-406 758-802	Reproductive (0.273) Gastrointestinal (0.227) Cardiovascular (0.136) Musculoskeletal (0.136)	Cancer (0.681) Cell Proliferation (0.182) Inflammation/Trauma (0.318)	pINCY
89	272-316	Reproductive (0.229) Gastrointestinal (0.193) Nervous (0.193)	Cancer (0.404) Inflammation (0.220) Cell Proliferation (0.165)	pINCY

Table 3 (cont.)

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
90	98-142	Nervous (0.400) Cardiovascular (0.200) Developmental (0.200) Gastrointestinal (0.200)	Cell Proliferation (0.400) Inflammation (0.400) Cancer (0.200)	pINCY
91	384-428 2016-2060	Reproductive (0.221) Gastrointestinal (0.156) Hematopoietic/Immune (0.143)	Cancer (0.468) Inflammation/Trauma (0.325) Cell Proliferation (0.273)	PBLUESCRIPT
92	80-124 731-775	Reproductive (0.286) Hematopoietic/Immune (0.143) Nervous (0.143)	Cancer (0.469) Inflammation/Trauma (0.326) Cell Proliferation (0.306)	PBLUESCRIPT
93	437-481 641-685	Reproductive (0.250) Nervous (0.200) Cardiovascular (0.183)	Cancer (0.550) Inflammation/Trauma (0.284) Cell Proliferation (0.150)	PBLUESCRIPT
94	397-441 1036-1080	Reproductive (0.291) Hematopoietic/Immune (0.228) Nervous (0.152)	Inflammation/Trauma (0.468) Cancer (0.392) Cell Proliferation (0.165)	pINCY
95	247-291	Reproductive (0.242) Hematopoietic/Immune (0.121) Nervous (0.121) Urologic (0.121)	Cancer (0.455) Inflammation/Trauma (0.333) Cell Proliferation (0.273)	pINCY
96	453-497 858-902	Nervous (0.600) Reproductive (0.400)	Cancer (0.400) Inflammation/Trauma (0.200) Neurological (0.200)	pINCY
97	224-268 770-814 1211-1255	Gastrointestinal (0.262) Reproductive (0.215) Nervous (0.169)	Cancer (0.462) Inflammation/Trauma (0.339) Cell Proliferation (0.231)	pINCY
98	3-47 1086-1130	Reproductive (0.211) Gastrointestinal (0.211) Hematopoietic/Immune (0.158)	Cancer (0.553) Cell Proliferation (0.368) Inflammation/Trauma (0.342)	pINCY
99	388-432 874-918	Reproductive (0.268) Nervous (0.146) Cardiovascular (0.146)	Cancer (0.390) Inflammation/Trauma (0.390) Cell Proliferation (0.220)	pINCY
100	26-70	Gastrointestinal (0.238) Cardiovascular (0.190) Hematopoietic/Immune (0.143) Nervous (0.143) Endocrine (0.143)	Cancer (0.429) Inflammation/Trauma (0.381) Cell Proliferation (0.190)	pINCY

Table 3 (cont.)

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
101	226-270 2062-2106	Nervous (0.234) Hematopoietic/Immune (0.170) Reproductive (0.149)	Inflammation/Trauma (0.383) Cancer (0.362) Cell Proliferation (0.213)	pINCY
102	487-531	Reproductive (0.276) Nervous (0.161) Gastrointestinal (0.138) Cardiovascular (0.138)	Cancer (0.494) Cell Proliferation (0.310) Inflammation/Trauma (0.264)	pINCY
103	561-605	Reproductive (0.274) Gastrointestinal (0.194) Cardiovascular (0.129)	Cancer (0.452) Inflammation/Trauma (0.339) Cell Proliferation (0.258)	pINCY
104	287-331 806-850	Gastrointestinal (0.500) Reproductive (0.250) Musculoskeletal (0.250)	Cancer (0.500) Inflammation/Trauma (0.250)	pINCY
105	154-198 505-549 757-801	Gastrointestinal (0.233) Reproductive (0.209) Hematopoietic/Immune (0.163) Nervous (0.163)	Cancer (0.465) Inflammation/Trauma (0.326) Cell Proliferation (0.209)	pINCY
106	174-218 1182-1226	Reproductive (0.185) Hematopoietic/Immune (0.185) Nervous (0.185)	Inflammation/Trauma (0.352) Cell Proliferation (0.333) Cancer (0.315)	pINCY
107	120-164 489-533	Reproductive (0.231) Hematopoietic/Immune (0.231) Nervous (0.154) Cardiovascular (0.154)	Cell Proliferation (0.462) Inflammation/Trauma (0.385) Cancer (0.231)	pINCY
108	64-108 1738-1782	Nervous (0.277) Reproductive (0.255) Cardiovascular (0.160)	Cancer (0.362) Inflammation/Trauma (0.362) Cell Proliferation (0.149)	pINCY
109	415-459 1027-1071 1549-1593	Reproductive (0.274) Hematopoietic/Immune (0.226) Nervous (0.167)	Inflammation/Trauma (0.476) Cancer (0.393) Cell Proliferation (0.179)	pINCY
110	242-286	Reproductive (0.500) Nervous (0.500)	Cancer (1.000)	pINCY
111	488-541 1028-1081	Reproductive (0.270) Nervous (0.191) Gastrointestinal (0.126)	Cancer (0.507) Inflammation/Trauma (0.284) Cell Proliferation (0.172)	PSPORT1

Table 3 (cont.)

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
112	273-326 867-920 1299-1352	Reproductive (0.312) Nervous (0.281) Gastrointestinal (0.094)	Cancer (0.469) Inflammation/Trauma (0.328) Cell Proliferation (0.172)	pINCY
113	866-1135	Reproductive (0.245) Gastrointestinal (0.136) Nervous (0.136)	Cancer (0.445) Cell Proliferation (0.227) Inflammation/Trauma (0.327)	pINCY
114	155-325 812-1105	Nervous (0.314) Reproductive (0.275) Gastrointestinal (0.098)	Cancer (0.471) Inflammation/Trauma (0.118)	pINCY
115	14-298	Gastrointestinal (0.190) Nervous (0.190) Reproductive (0.190)	Cancer (0.476) Cell Proliferation (0.190) Inflammation/Trauma (0.238)	pINCY
116	41-235	Reproductive (0.400) Nervous (0.267) Musculoskeletal (0.133)	Cancer (0.600) Inflammation/Trauma (0.334) Cell Proliferation (0.067)	PSPORT1
117	379-432 973-1026 1297-1350	Reproductive (0.327) Nervous (0.184) Urologic (0.102)	Cancer (0.531) Cell Proliferation (0.224) Inflammation/Trauma (0.265)	pINCY
118	974-1465	Reproductive (0.231) Nervous (0.190) Gastrointestinal (0.169)	Cancer (0.446) Inflammation/Trauma (0.343) Cell Proliferation (0.226)	pINCY
119	543-1028	Reproductive (0.292) Nervous (0.163) Gastrointestinal (0.139)	Cancer (0.517) Cell Proliferation (0.167) Inflammation/Trauma (0.235)	PSPORT1
120	385-552	Nervous (0.571) Cardiovascular (0.143) Developmental (0.143)	Cancer (0.429) Inflammation/Trauma (0.572) Cell Proliferation (0.143)	pINCY
121	685-864	Nervous (0.300) Hematopoietic/Immune (0.200) Cardiovascular (0.140)	Cancer (0.340) Inflammation/Trauma (0.440) Cell Proliferation (0.200)	pINCY
122	703-1026	Reproductive (0.400) Cardiovascular (0.160) Nervous (0.160)	Cancer (0.680) Cell Proliferation (0.120) Inflammation/Trauma (0.160)	pINCY
123	830-1351	Reproductive (0.200) Cardiovascular (0.154) Hematopoietic/Immune (0.154)	Cancer (0.415) Cell Proliferation (0.277) Inflammation/Trauma (0.354)	pINCY

Table 3 (cont.)

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
124	272-325	Cardiovascular (0.250) Gastrointestinal (0.250) Musculoskeletal (0.250)	Inflammation/Trauma (0.750)	pINCY
125	130-972	Reproductive (0.180) Cardiovascular (0.160) Hematopoietic/Immune (0.160)	Cancer (0.440) Inflammation/Trauma (0.340) Cell Proliferation (0.220)	pINCY
126	434-973	Reproductive (0.188) Cardiovascular (0.156) Gastrointestinal (0.156)	Cancer (0.422) Inflammation/Trauma (0.328) Cell Proliferation (0.203)	pINCY
127	489-899	Gastrointestinal (0.333) Reproductive (0.333) Nervous (0.125)	Cancer (0.625) Inflammation/Trauma (0.208) Cell Proliferation (0.042)	pINCY
128	19-1242	Reproductive (0.354) Nervous (0.188) Gastrointestinal (0.146)	Cancer (0.562) Cell Proliferation (0.250) Inflammation/Trauma (0.250)	pINCY
129	217-270 541-594	Reproductive (0.364) Cardiovascular (0.182) Gastrointestinal (0.182)	Cancer (0.636) Inflammation/Trauma (0.364)	pINCY
130	115-864	Gastrointestinal (0.250) Hematopoietic/Immune (0.208) Nervous (0.208)	Cancer (0.500) Inflammation/Trauma (0.292)	pINCY
131	255-308	Reproductive (0.265) Nervous (0.169) Gastrointestinal (0.120)	Cancer (0.482) Cell Proliferation (0.349) Inflammation/Trauma (0.253)	pINCY
132	23-541	Nervous (0.909) Endocrine (0.091)	Cancer (0.636) Cell Proliferation (0.091) Inflammation/Trauma (0.182)	pINCY

Table 4

SEQ ID NO:	Library	Library Comment
67	LATRTUT02	Library was constructed using RNA isolated from a myxoma removed from the left atrium of a 43-year-old Caucasian male during annuloplasty. Pathology indicated atrial myxoma. Patient history included pulmonary insufficiency, acute myocardial infarction, atherosclerotic coronary artery disease, and hyperlipidemia. Family history included benign hypertension, acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
68	PENITUT01	Library was constructed using RNA isolated from tumor tissue removed from the penis of a 64-year-old Caucasian male during penile amputation. Pathology indicated a fungating invasive grade 4 squamous cell carcinoma involving the inner wall of the foreskin and extending onto the glans penis. Patient history included benign neoplasm of the large bowel, atherosclerotic coronary artery disease, angina pectoris, gout, and obesity. Family history included malignant pharyngeal neoplasm, chronic lymphocytic leukemia, and chronic liver disease.
69	BLADTUT04	Library was constructed using RNA isolated from bladder tumor tissue removed from a 60-year-old Caucasian male during a radical cystectomy, prostatectomy, and vasectomy. Pathology indicated grade 3 transitional cell carcinoma in the left bladder wall. Carcinoma in-situ was identified in the dome and trigone. Family history included type I diabetes, a malignant neoplasm of the stomach, atherosclerotic coronary artery disease, and an acute myocardial infarction.
70	BLADTUT06	Library was constructed using RNA isolated from bladder tumor tissue removed from the posterior bladder wall of a 58-year-old Caucasian male during a radical cystectomy, radical prostatectomy, and gastrostomy. Pathology indicated grade 3 transitional cell carcinoma in the left lateral bladder wall. The remaining bladder showed marked cystitis with scattered microscopic foci of transitional cell carcinoma in situ. Patient history included angina and emphysema. Family history included acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
71	ADRENOT07	Library was constructed using RNA isolated from adrenal tissue removed from a 61-year-old female during a bilateral adrenalectomy. Patient history included an unspecified disorder of the adrenal glands.
72	BRSTNOT19	Library was constructed using RNA isolated from breast tissue removed from a 67-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated residual invasive lobular carcinoma. Patient history included depressive disorder, benign large bowel neoplasm, and hemorrhoids. Family history included cerebrovascular and cardiovascular disease and lung cancer.

Table 4 (cont.)

SEQ ID NO:	Library	Library Comment
73	SPLNNOT12	Library was constructed using RNA isolated from spleen tissue removed from a 65-year-old female. Pathology indicated the spleen was negative for metastasis. Pathology for the associated tumor tissue indicated well-differentiated neuroendocrine carcinoma (islet cell tumor), nuclear grade 1, forming a dominant mass in the distal pancreas. Multiple smaller tumor nodules were immediately adjacent to the main mass. The liver showed metastatic grade 1 islet cell tumor, forming multiple nodules. Multiple (4) pericholedochal lymph nodes contained metastatic grade 1 islet cell tumor.
74	MONOTXT02	Library was constructed using RNA isolated from treated monocytes from peripheral blood removed from a 42-year-old female. The cells were treated with interleukin-10 (IL-10) and lipopolysaccharide (LPS). IL-10 was added at time 0 at 10 ng/ml, LPS was added at 1 hour at 5 ng/ml. The monocytes were isolated from buffy coat by adherence to plastic. Incubation time was 24 hours.
75	FIBPFEN06	Library was constructed from 1.56 million independent clones from a prostate stromal fibroblast tissue library. Starting RNA was made from fibroblasts of prostate stroma removed from a male fetus,who died after 26 weeks' gestation. The libraries were normalized in two rounds using conditions adapted from Soares et al. (1994) Proc. Natl. Acad. Sci. USA 91:9228 and Bonaldo et al.(1996) Genome Research 6:791, except that a significantly longer (48-hours/round) reannealing hybridization was used.
76	HUVESTB01	Library was constructed using RNA isolated from shear-stressed HUV-EC-C (ATCC CRL 1730) cells. Before RNA isolation, the cells were subjected to a shear stress of 10 dynes/cm.
77	SYNOOAT01	Library was constructed using RNA isolated from the knee synovial membrane tissue of an 82-year-old female with osteoarthritis.
78	UTRSNOT05	Library was constructed using RNA isolated from the uterine tissue of a 45-year-old Caucasian female during a total abdominal hysterectomy and total colectomy. Pathology for the associated tumor tissue indicated multiple leiomyomas of the myometrium and a grade 2 colonic adenocarcinoma of the cecum. Patient history included multiple sclerosis and mitral valve disorder. Family history included type I diabetes, cerebrovascular disease, atherosclerotic coronary artery disease, malignant skin neoplasm, hypertension, and malignant neoplasm of the colon.
79	HIPONON01	Library was constructed from 1.13 million independent clones from a hippocampus library. RNA was isolated from the hippocampus tissue of a 72-year-old Caucasian female who died from an intracranial bleed. Patient history included nose cancer, hypertension, and arthritis. The normalization and hybridization conditions were adapted from Soares et al. (1994) Proc. Natl. Acad. Sci. USA 91:9228.

Table 4 (cont.)

SEQ ID NO:	Library	Library Comment
80	BRSTTUT03	Library was constructed using RNA isolated from breast tumor tissue removed from a 58-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated multicentric invasive grade 4 lobular carcinoma. The mass was identified in the upper outer quadrant, and three separate nodules were found in the lower outer quadrant of the left breast. Patient history included skin cancer, rheumatic heart disease, osteoarthritis, and tuberculosis. Family history included cerebrovascular disease, coronary artery aneurysm, breast cancer, prostate cancer, atherosclerotic coronary artery disease, and type I diabetes.
81	SININOT01	Library was constructed using RNA isolated from ileum tissue obtained from the small intestine of a 4-year-old Caucasian female, who died from a closed head injury. Patient history included jaundice. Previous surgeries included a double hernia repair.
82	SINTFET03	Library was constructed using RNA isolated from small intestine tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation.
83	HNT3AZT01	Library was constructed using RNA isolated from the hNT2 cell line (derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor). Cells were treated for three days with 0.35 micromolar 5-aza-2'-deoxycytidine (AZ).
84	ENDANOT01	Library was constructed using RNA isolated from aortic endothelial cell tissue from an explanted heart removed from a male during a heart transplant.
85	LUNGTUT08	Library was constructed using RNA isolated from lung tumor tissue removed from a 63-year-old Caucasian male during a right upper lobectomy with fiberoptic bronchoscopy. Pathology indicated a grade 3 adenocarcinoma. Patient history included atherosclerotic coronary artery disease, an acute myocardial infarction, rectal cancer, an asymtomatic abdominal aortic aneurysm, tobacco abuse, and cardiac dysrhythmia. Family history included congestive heart failure, stomach cancer, and lung cancer, type II diabetes, atherosclerotic coronary artery disease, and an acute myocardial infarction.
86	OVARTUT10	Library was constructed using RNA isolated from ovarian tumor tissue removed from the left ovary of a 58-year-old Caucasian female during a total abdominal hysterectomy, removal of a solitary ovary, and repair of inguinal hernia. Pathology indicated a metastatic grade 3 adenocarcinoma of colonic origin, forming a partially cystic and necrotic tumor mass in the left ovary, and an adenocarcinoma of colonic origin, forming a nodule in the left mesovarium. A single intramural leiomyoma was identified in the myometrium. The cervix showed mild chronic cystic cervicitis. Patient history included benign hypertension, follicular cyst of the ovary, colon cancer, benign colon neoplasm, and osteoarthritis. Family history included emphysema, myocardial infarction, atherosclerotic coronary artery disease, benign hypertension, and hyperlipidemia.

Table 4 (cont.)

SEQ ID NO:	Library	Library Comment
87	BRSTNOT13	Library was constructed using RNA isolated from breast tissue removed from a 36-year-old Caucasian female during bilateral simple mastectomy. Patient history included a breast neoplasm, depressive disorder, hyperlipidemia, and a chronic stomach ulcer. Family history included cardiovascular and cerebrovascular disease; hyperlipidemia; skin, breast, esophageal, bladder, and bone cancer; and Hodgkin's lymphoma.
88	UTRSNOR01	Library was constructed using RNA isolated from uterine endometrium tissue removed from a 29-year-old Caucasian female during a vaginal hysterectomy and cystocele repair. Pathology indicated the endometrium was secretory, and the cervix showed mild chronic cervicitis with focal squamous metaplasia. Pathology for the associated tumor tissue indicated intramural uterine leiomyoma. Patient history included hypothyroidism, pelvic floor relaxation, and paraplegia. Family history included benign hypertension, type II diabetes, and hyperlipidemia.
89	BRSTTMT02	Library was constructed using RNA isolated from diseased right breast tissue removed from a 46-year-old Caucasian female during a unilateral extended simple mastectomy and open breast biopsy. Pathology indicated mildly proliferative fibrocystic change, including intraductal duct ectasia, papilloma formation, and ductal hyperplasia. Pathology for the associated tumor tissue indicated multifocal ductal carcinoma in situ, both comedo and non-comedo types, nuclear grade 2 with extensive intraductal calcifications. Patient history included deficiency anemia, normal delivery, chronic sinusitis, extrinsic asthma, and kidney infection. Family history included type II diabetes, benign hypertension,cerebrovascular disease, skin cancer, and hyperlipidemia.
90	LIVRDIR01	Library was constructed using RNA isolated from diseased liver tissue removed from a 63-year-old Caucasian female during a liver transplant.Patient history included primary biliary cirrhosis. Serology was positive for anti-mitochondrial antibody.
91	HUVENOB01	Library was constructed using RNA isolated from HUV-EC-C (ATCC CRL 1730) cells.
92	TESTNOT03	Library was constructed using RNA isolated from testicular tissue removed from a 37-year-old Caucasian male, who died from liver disease. Patient history included cirrhosis, jaundice, and liver failure.
93	LUNGNOT02	Library was constructed using RNA isolated from the lung tissue of a 47-year-old Caucasian male, who died of a subarachnoid hemorrhage.
94	LUNGFET03	Library was constructed using RNA isolated from lung tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation.
95	PANCNOT07	Library was constructed using RNA isolated from the pancreatic tissue of a Caucasian male fetus, who died at 23 weeks' gestation.

Table 4 (cont.)

SEQ ID NO:	Library	Library Comment
96	BRAINOT12	Library was constructed using RNA isolated from brain tissue removed from the right frontal lobe of a 5-year-old Caucasian male during a hemispherectomy. Pathology indicated extensive polymicrogyria and mild to moderate gliosis (predominantly subpial and subcortical), which are consistent with chronic seizure disorder. Family history included a cervical neoplasm.
97	LIVRTUT01	Library was constructed using RNA isolated from liver tumor tissue removed from a 51-year-old Caucasian female during a hepatic lobectomy. Pathology indicated metastatic grade 3 adenocarcinoma consistent with colon cancer. Family history included a malignant neoplasm of the liver.
98	GBLATUT01	Library was constructed using RNA isolated from gall bladder tumor tissue removed from a 78-year-old Caucasian female during a cholecystectomy. Pathology indicated invasive grade 2 squamous cell carcinoma, forming a mass in the gall bladder. Patient history included diverticulitis of the colon, palpitations, benign hypertension, and hyperlipidemia. Family history included a cholecystectomy, atherosclerotic coronary artery disease, hyperlipidemia, and benign hypertension.
99	LEUKNOT02	Library was constructed using RNA isolated from white blood cells of a 45-year-old female with blood type O+. The donor tested positive for cytomegalovirus (CMV).
100	LUNGNOT22	Library was constructed using RNA isolated from lung tissue removed from a 58-year-old Caucasian female. The tissue sample used to construct this library was found to have tumor contaminant upon microscopic examination. Pathology for the associated tumor tissue indicated a caseating granuloma. Family history included congestive heart failure, breast cancer, secondary bone cancer, acute myocardial infarction and atherosclerotic coronary artery disease.
101	ADRETUT06	Library was constructed using RNA isolated from adrenal tumor tissue removed from a 57-year-old Caucasian female during a unilateral right adrenalectomy. Pathology indicated pheochromocytoma, forming a nodular mass completely replacing the medulla of the adrenal gland.
102	ADRETUT06	Library was constructed using RNA isolated from adrenal tumor tissue removed from a 57-year-old Caucasian female during a unilateral right adrenalectomy. Pathology indicated pheochromocytoma, forming a nodular mass completely replacing the medulla of the adrenal gland.
103	THYRNOT10	Library was constructed using RNA isolated from diseased left thyroid tissue removed from a 30-year-old Caucasian female during a unilateral thyroid lobectomy and parathyroid reimplantation. Pathology indicated lymphocytic thyroiditis.

Table 4 (cont.)

SEQ ID NO:	Library	Library Comment
104	CONNTUT05	Library was constructed using RNA isolated from tumorous skull soft tissue removed from a 34-year-old Caucasian female during skull lesion excision. Pathology indicated grade 3 ependymoma forming an implant in the dermis and subcutis associated with dense fibrosis. Patient history included seizures, bone cancer, and brain cancer. Surgeries included cranioplasty and cerebral meninges lesion excision, and treatment included whole brain radiation. Family history included anxiety and depression.
105	HEAANOT01	Library was constructed using RNA isolated from right coronary and right circumflex coronary artery tissue removed from the explanted heart of a 46-year-old Caucasian male during a heart transplantation. Patient history included myocardial infarction from total occlusion of the left anterior descending coronary artery, atherosclerotic coronary artery disease, hyperlipidemia, myocardial ischemia, dilated cardiomyopathy, left ventricular dysfunction, and tobacco abuse. Family history included atherosclerotic coronary artery disease.
106	UTRMTMT01	Library was constructed using RNA isolated from myometrial tissue removed from a 45-year-old Caucasian female during vaginal hysterectomy and bilateral salpingo-oophorectomy. Pathology indicated the myometrium was negative for tumor. Pathology for the associated tumor tissue indicated multiple (23) subserosal, intramural, and submucosal leiomyomata. The endometrium was in proliferative phase. The right ovary contained an old corpus luteum. The cervix, left ovary, and right and left fallopian tubes were unremarkable. The patient presented with stress incontinence. Patient history included extrinsic asthma without status asthmaticus and normal delivery. Patient medications included Motrin, iron sulfate, Premarin, prednisone, Tylenol #3, and Colace. Family history included cerebrovascular disease, depression, and atherosclerotic coronary artery disease.
107	FIBPFEN06	This normalized library was constructed from 1.56 million independent clones from a prostate stromal fibroblast library. RNA was isolated from a male fetus, who died after 26 weeks' gestation. The normalization and hybridization conditions were adapted from Soares et al. (1994) Proc. Natl. Acad. Sci. USA 91:9228.

Table 4 (cont.)

SEQ ID NO:	Library	Library Comment
108	BRAINOT19	Library was constructed using RNA isolated from diseased brain tissue removed from the left frontal lobe of a 27-year-old Caucasian male during a brain lobectomy. Pathology indicated a focal deep white matter lesion, characterized by marked gliosis, calcifications, and hemosiderin-laden macrophages, consistent with a remote perinatal injury. This tissue also showed mild to moderate generalized gliosis, predominantly subpial and subcortical, consistent with chronic seizure disorder. The left temporal lobe, including the mesial temporal structures, showed focal, marked pyramidal cell loss and gliosis in hippocampal sector CA1, consistent with mesial temporal sclerosis. GFAP was positive for astrocytes. Patient presented with intractable epilepsy, focal epilepsy, hemiplegia, and an unspecified brain injury. Patient history included cerebral palsy, abnormality of gait, and depressive disorder. Family history included brain cancer.
109	COLCDIT03	Library was constructed using RNA isolated from diseased colon polyp tissue removed from the cecum of a 67-year-old female. Pathology indicated a benign cecum polyp. Pathology for the associated tumor tissue indicated invasive grade 3 adenocarcinoma that arose in tubulovillous adenoma forming a fungating mass in the cecum.
110	BRAXNOT03	Library was constructed using RNA isolated from sensory-motor cortex tissue removed from the brain of a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. The cerebral hemisphere revealed moderate fibrosis of the leptomeninges with focal calcifications. There was evidence of shrunken and slightly eosinophilic pyramidal neurons throughout the cerebral hemispheres. There were also multiple small microscopic areas of cavitation with surrounding gliosis, scattered throughout the cerebral cortex. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly and an enlarged spleen and liver. Patient medications included simethicone, Lasix, Digoxin, Colace, Zantac, Captopril, and Vasotec.
111	BRAITUT02	Library was constructed using RNA isolated from brain tumor tissue removed from the frontal lobe of a 58-year-old Caucasian male during excision of a cerebral meningeal lesion. Pathology indicated a grade 2 metastatic hypernephroma. Patient history included a grade 2 renal cell carcinoma, insomnia, and chronic airway obstruction. Family history included a malignant neoplasm of the kidney.
112	PROSNOT11	Library was constructed using RNA isolated from the prostate tissue of a 28-year-old Caucasian male, who died from a self-inflicted gunshot wound.
113	LIVRTUT01	Library was constructed using RNA isolated from liver tumor tissue removed from a 51-year-old Caucasian female during a hepatic lobectomy. Pathology indicated metastatic grade 3 adenocarcinoma consistent with colon cancer. Family history included a malignant neoplasm of the liver.

Table 4 (cont.)

SEQ ID NO:	Library	Library Comment
114	PANCTUT02	Library was constructed using RNA isolated from pancreatic tumor tissue removed from a 45-year-old Caucasian female during radical pancreaticoduodenectomy. Pathology indicated a grade 4 anaplastic carcinoma. Family history included benign hypertension, hyperlipidemia and atherosclerotic coronary artery disease.
115	LIVRFET02	Library was constructed using RNA isolated from liver tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation. Family history included seven days of erythromycin treatment for bronchitis in the mother during the first trimester.
116	BRAITUT03	Library was constructed using RNA isolated from brain tumor tissue removed from the left frontal lobe a 17-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated a grade 4 fibrillary giant and small-cell astrocytoma. Family history included benign hypertension and cerebrovascular disease.
117	BRSTNOT07	Library was constructed using RNA isolated from diseased breast tissue removed from a 43-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated mildly proliferative fibrocystic changes with epithelial hyperplasia, papillomatosis, and duct ectasia. Pathology for the associated tumor tissue indicated invasive grade 4, nuclear grade 3 mammary adenocarcinoma with extensive comedo necrosis. Family history included epilepsy, cardiovascular disease, and type II diabetes.
118	SMCANOT01	Library was constructed using RNA isolated from an aortic smooth muscle cell line derived from the explanted heart of a male during a heart transplant.
119	THP1AZS08	Library was constructed using 5.76 million clones from a 5-aza-2'-deoxycytidine (AZ) treated THP-1 promonocyte cell line library. Starting RNA was made from THP-1 promonocyte cells treated for three days with 0.8 micromolar AZ. The hybridization probe for subtraction was derived from a similarly constructed library, made from 1 microgram of polyA RNA isolated from untreated THP-1 cells. 5.76 million clones from the AZ-treated THP-1 cell library were then subjected to two rounds of subtractive hybridization with 5 million clones from the untreated THP-1 cell library. Subtractive hybridization conditions were based on the methodologies of Swaroop et al. (1991) Nucleic Acids Res. 19:1954, and Bonaldo et al. (1996) Genome Research 6:791. THP-1 (ATCC TIB 202) is a human promonocyte line derived from peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia (ref: Int. J. Cancer (1980) 26:171).
120	ADRETUT06	Library was constructed using RNA isolated from adrenal tumor tissue removed from a 57-year-old Caucasian female during a unilateral right adrenalectomy. Pathology indicated pheochromocytoma, forming a nodular mass completely replacing the medulla of the adrenal gland.

Table 4 (cont.)

SEQ ID NO:	Library	Library Comment
121	SININOT03	Library was constructed using RNA isolated from ileum tissue obtained from an 8-year-old Caucasian female, who died from head trauma. Serology was positive for cytomegalovirus (CMV).
122	SININOT03	Library was constructed using RNA isolated from ileum tissue obtained from an 8-year-old Caucasian female, who died from head trauma. Serology was positive for cytomegalovirus (CMV).
123	TLYMNOT06	Library was constructed using RNA isolated from activated Th2 cells. These cells were differentiated from umbilical cord CD4 T cells with IL-4 in the presence of anti-IL-12 antibodies and B7-transfected COS cells, and then activated for six hours with anti-CD3 and anti-CD28 antibodies.
124	HEAANOT01	Library was constructed using RNA isolated from right coronary and right circumflex coronary artery tissue removed from the explanted heart of a 46-year-old Caucasian male during a heart transplantation. Patient history included myocardial infarction from total occlusion of the left anterior descending coronary artery, atherosclerotic coronary artery disease, hyperlipidemia, myocardial ischemia, dilated cardiomyopathy, left ventricular dysfunction, and tobacco abuse. Previous surgeries included cardiac catheterization. Family history included atherosclerotic coronary artery disease.
125	TLYJINT01	Library was constructed using RNA isolated from a Jurkat cell line derived from the T cells of a male. The cells were treated for 18 hours with 50 ng/ml phorbol ester (PMA) and 1 micromolar calcium ionophore. Patient history included acute T-cell leukemia.
126	BRAITUT24	Library was constructed using RNA isolated from right frontal brain tumor tissue removed from a 50-year-old Caucasian male during a cerebral meninges lesion excision. Pathology indicated meningioma. Family history included colon cancer and cerebrovascular disease.
127	PROSTUT16	Library was constructed using RNA isolated from prostate tumor tissue removed from a 55-year-old Caucasian male. Pathology indicated adenocarcinoma, Gleason grade 5+4. Adenofibromatous hyperplasia was also present. The patient presented with elevated prostate specific antigen (PSA). Patient history included calculus of the kidney. Family history included lung cancer and breast cancer.
128	BRONNOT01	Library was constructed using RNA isolated from bronchial tissue removed from a 15-year-old Caucasian male.

Table 4 (cont.)

SEQ ID NO:	Library	Library Comment
129	BLADTUT03	Library was constructed using RNA isolated from bladder tumor tissue removed from a 58-year-old Caucasian male during a radical cystectomy, radical prostatectomy, regional lymph node excision, and urinary diversion to bowel. Pathology indicated invasive grade 3 transitional cell carcinoma. Patient history included a benign colon neoplasm. Family history included cerebrovascular disease and atherosclerotic coronary artery disease.
130	COLXTDT01	Library was constructed using RNA isolated from colon tissue removed from the appendix of a 37-year-old Black female during myomectomy, dilation and curettage, right fimbrial region biopsy, and incidental appendectomy. Pathology indicated an unremarkable appendix. Pathology for the associated tumor tissue indicated multiple (12) uterine leiomyomata. Patient history included premenopausal menorrhagia and sarcoidosis of the lung. Family history included acute myocardial infarction and atherosclerotic coronary artery disease.
131	BRATNOT02	Library was constructed using RNA isolated from superior temporal cortex tissue removed from the brain of a 35-year-old Caucasian male. No neuropathology was found. Patient history included dilated cardiomyopathy, congestive heart failure, and an enlarged spleen and liver.
132	BRAWNOT01	Library was constructed using RNA isolated from dentate nucleus tissue removed from the brain of a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly, and an enlarged spleen and liver.

Table 5 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

SUBSTITUTE SHEET (RULE 26)

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

a) an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, and SEQ ID NO:66,

b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, and SEQ ID NO:66,

c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:52, SEQ ID

NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, and SEQ ID NO:66, and

d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, and SEQ ID NO:66.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, and SEQ ID NO:66.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID

NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, and SEQ ID NO:132.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method for producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:118, SEQ ID NO:119,

SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, and SEQ ID NO:132,

b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, and SEQ ID NO:132,

c) a polynucleotide sequence complementary to a),

d) a polynucleotide sequence complementary to b), and

e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, and SEQ ID NO:132.

18. A method for treating a disease or condition associated with decreased expression of functional GBAP, comprising administering to a patient in need of such treatment the pharmaceutical composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional GBAP, comprising administering to a patient in need of such treatment a pharmaceutical

composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional GBAP, comprising administering to a patient in need of such treatment a pharmaceutical composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, and

b) detecting altered expression of the target polynucleotide.

28. A method for assessing toxicity of a test compound, said method comprising:

a) treating a biological sample containing nucleic acids with the test compound;

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;

c) quantifying the amount of hybridization complex; and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.